

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2004, 02:22:49 ; Search time 2482 Seconds
(without alignments)
10076.120 Million cell updates/sec

Title: US-09-867-753-1

Perfect score: 577

Sequence: 1 tccacatcagcgcgtccag.....tctacatcgcgtgagactag 577

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ha: *
2: gb_hcg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_scs: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_scs: *
28: em_un: *
29: em_vl: *
30: em_hcg_hum: *
31: em_hcg_inv: *
32: em_hcg_other: *
33: em_hcg_mus: *
34: em_hcg_pln: *
35: em_hcg_rod: *
36: em_hcg_mam: *
37: em_hcg_vrt: *
38: em_ay: *
39: em_hngo_hum: *
40: em_hngo_mus: *
41: em_hngo_other: *

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	577	100.0	577	6	AX322486	AX322486 Sequence
2	566	98.1	755	6	AY099086	AY099086 Homo sapi
3	566	98.1	979	6	AR339190	AR339190 Sequence
C 4	409	70.9	10968	6	AX322488	AX322488 Sequence
C 5	409	70.9	170967	9	AC005023	AC005023 Homo sapi
C 6	399.4	69.2	139483	2	AC145687	AC145687 Pan trogl
C 7	383.4	66.4	204520	2	AC146356	AC146356 Pongo pyg
8	348.2	60.3	157119	2	AC146353	AC146353 Macaca mu
9	106.8	18.5	1205	9	HS421131	AL590524 Novel hum
10	105.8	18.3	1194	9	AK058125	AK058125 Homo sapi
11	105.8	18.3	1219	9	BC021719	BC021719 Homo sapi
12	105.8	18.3	1274	9	AF317219	AF317219 Homo sapi
13	105.8	18.3	1297	9	HS525N141	AL590526 Novel hum
14	103	17.9	1451	9	AY114148	AY114148 Homo sapi
15	103	17.9	1518	6	AX834967	AX834967 Sequence
16	103	17.9	1518	9	AK097704	AK097704 Homo sapi
17	103	17.9	1580	9	BC053599	BC053599 Homo sapi
C 18	102	17.7	157119	2	AC146353	AC146353 Macaca mu
19	97.2	16.8	1325	10	AF017734	AF017734 Mus muscu
20	97.2	16.8	1408	10	AF017735	AF017735 Mus muscu
21	97.2	16.8	1616	10	AF085715	AF085715 Mus muscu
22	97.2	16.8	1720	10	AF004211	AF004211 Mus muscu
23	90.6	15.7	175366	10	AC101718	AC101718 Mus muscu
24	74.8	13.0	4329	5	AF001393	AF001393 Oryzias l
25	73.8	12.8	1764	5	XLAFO01049	AF001049 Xenopus l
26	73	12.7	868	10	BC049711	BC049711 Mus muscu
27	72.4	12.5	1913	3	AF165886	AF165886 Branchio
28	71.8	12.4	882	10	AF017453	AF017453 Mus muscu
29	71.8	12.4	1326	9	BC010923	BC010923 Homo sapi
30	71.8	12.4	1414	5	AB006104	AB006104 Danio rer
31	71.8	12.4	1442	9	HSU31986	U131986 Human carti
32	71.6	12.4	1719	5	AY130460	AY130460 Xenopus l
33	71.6	12.4	220851	2	AC097956	AC097956 Rattus no
C 34	71.2	12.3	286	10	AY116506	AY116506 Mus muscu
35	70.2	12.2	2368	5	XLAFO01048	AF001048 Xenopus l
36	70	12.1	740	10	AY147207	AY147207 Mus muscu
37	70	12.1	1639	5	BC056324	BC056324 Danio rer
38	69.6	12.1	624	10	S82627	S82627 Rattus sp.
39	69.6	12.1	1494	5	AF121771	AF121771 Danio rer
40	69.2	12.0	1164	5	ZERGHDA	L03394 Brachydanio
41	69.2	12.0	1252	5	ZERGHDA	L03395 Brachydanio
42	68.6	11.9	1485	5	AF017273	AF017273 Xenopus l
43	68	11.8	1164	5	AF092538	AF092538 Gallus ga
44	68	11.8	2399	10	RN029174	U29174 Rattus norv
45	67.6	11.7	2032	5	PMA456324	AJ458324 Petromyzo

ALIGNMENTS

RESULT 1
AX322486
LOCUS Sequence 1 from Patent EP1162265.
DEFINITION AX322486
ACCESSION AX322486
VERSION AX322486.1 GI:18093553
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Haendler, B., Weiss, B. and Geserick, C.
TITLE Human pem as a target for fertility control and alzheimer therapy
JOURNAL Patent: EP 1162265-A 1 12-DEC-2001;

Pred. No. is the number of results predicted by chance to have a

Schering Aktiengesellschaft (DE)

FEATURES
source

1..577
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 100.0%; Score 577; DB 6; Length 577;
Best Local Similarity 100.0%; Pred. No. 2e-113;
Matches 577; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACACATGAGGCGCTCCAGCCATGAGCGCGTTCGCTGCTCCAGACACCGTGTCTACTG 60
DB 1 TCACACATGAGGCGCTCCAGCCATGAGCGCGTTCGCTGCTCCAGACACCGTGTCTACTG 60
QY 61 CCTGAGTGTATACCAAGTAAAAATAAGCCCCACCTCAGCTGAGGGGCGCATCAAGCCG 120
DB 61 CCTGAGTGTATACCAAGTAAAAATAAGCCCCACCTCAGCTGAGGGGCGCATCAAGCCG 120
QY 121 AGAAGGCGCATGTTGGCCAGAGAGCTCCAGGCGCTCATGGGTAATATGAACCTTGAGGGCG 180
DB 121 AGAAGGCGCATGTTGGCCAGAGAGCTCCAGGCGCTCATGGGTAATATGAACCTTGAGGGCG 180
QY 181 TGTGAACCAAGCAAGCAAGCGCATGAACCGCATGAGCGGCGCATGATCCCGAGGCGCGGTG 240
DB 181 TGTGAACCAAGCAAGCAAGCGCATGAACCGCATGAGCGGCGCATGATCCCGAGGCGCGGTG 240
QY 241 AAAACCAAGGAGCTCCGCGACAGCGGCGGCGGCGGAGGAGCGGCGGCGGCGGCGCAT 300
DB 241 AAAACCAAGGAGCTCCGCGACAGCGGCGGCGGCGGAGGAGCGGCGGCGGCGGCGCAT 300
QY 301 GAGAGGTCGCGCAGCCCGAGAAACATGAGCGCACGAACTCGGCGCGACGAACTTCAGCTGT 360
DB 301 GAGAGGTCGCGCAGCCCGAGAAACATGAGCGCACGAACTCGGCGCGACGAACTTCAGCTGT 360
QY 361 GAGAGTGAAGAGAGCTGAAAGTGTTCGACACACTCAATACCTGATGTGCCAAG 420
DB 361 GAGAGTGAAGAGAGCTGAAAGTGTTCGACACACTCAATACCTGATGTGCCAAG 420
QY 421 AAGGGAACCTTGCAGAACTTGAAGTGTGACTGAAGCAAAAGTGGGTTGTTAAGAA 480
DB 421 AAGGGAACCTTGCAGAACTTGAAGTGTGACTGAAGCAAAAGTGGGTTGTTAAGAA 480
QY 481 TAAAGGGCCAGATGAGCGACATCAGAGAGAAATTAATGCTGCGCAATGAATAGTGC 540
DB 481 TAAAGGGCCAGATGAGCGACATCAGAGAGAAATTAATGCTGCGCAATGAATAGTGC 540
QY 541 TGAACCAAGCAAGTGTCTACTCATCGTGTGACTAG 577
DB 541 TGAACCAAGCAAGTGTCTACTCATCGTGTGACTAG 577

RESULT 2

AY099086

LOCUS Homo sapiens paired-like homeobox protein OTEX mRNA, complete cds.
DEFINITION Homo sapiens paired-like homeobox protein OTEX mRNA, complete cds.
ACCESSION AY099086
VERSION AY099086.1 GI:20563000

KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 755)
Geisler, C., Weiss, B., Schleuning, W.D. and Haendler, B.

of homeobox genes
OEX, an androgen-regulated human member of the paired-like class

JOURNAL Biochem. J. 366 (Pt 1), 367-375 (2002)

MEDLINE 22152918

PUBMED 11980563

REFERENCE 2 (bases 1 to 755)

AUTHORS Geserick, C., Weiss, B. and Haendler, B.

TITLE Direct Submission
JOURNAL Submitted (23-APR-2002) Experimental Oncology, Schering AG,
Muelletstr. 170-178, Berlin 13342, Germany

FEATURES
source

1..755
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="X"
/map="Xq24"
76..630
/codon_start=1
/product="paired-like homeobox protein OTEX"
/protein_id="AA022794.1"
/db_xref="GI:20563001"
/translation="MARSLVHDTVRYCLSVYQVKISPTPOLGAASSAEHVGAGL
MGNMNPBGVNHENMNRDGMIPBGGGNBPPOPOPPPEPQAAAEHGPENMO
PRTRKTKTLLOVELEBSVFRHTQYPDVPTRELEBNLGVTEDKYRWFKNRACRR
HORELMLENLRAEDPDVCYIVD"

CDS

Query Match 98.1%; Score 566; DB 9; Length 755;
Best Local Similarity 100.0%; Pred. No. 4.5e-111;
Matches 566; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN

QY 12 GCGCTCCAGCCATGAGCGCGTTCGCTGCTCCAGACACCGTGTCTACTGAGTGTAT 71
DB 65 GCGCTCCAGCCATGAGCGCGTTCGCTGCTCCAGACACCGTGTCTACTGAGTGTAT 71
QY 72 ACCAGGTAAATAATAGCCCAACCTCAGCTGAGGGGCGAGCATCAAGCGAAGAGCGATG 131
DB 125 ACCAGGTAAATAATAGCCCAACCTCAGCTGAGGGGCGAGCATCAAGCGAAGAGCGATG 184
QY 132 TTGGCCAAAGAGCTCCAGGCGCTCATGGGTAATATGAACCTTGAGGGCGGTGTGAACCA 191
DB 185 TTGGCCAAAGAGCTCCAGGCGCTCATGGGTAATATGAACCTTGAGGGCGGTGTGAACCA 244
QY 192 AGAAGCGCATGAACCGCGATGAGCGGCGATATCCCGAGGGCGCGGTGGAACCAAGAGC 251
DB 245 AGAAGCGCATGAACCGCGATGAGCGGCGATATCCCGAGGGCGCGGTGGAACCAAGAGC 304
QY 252 CTGCGAGAGAGCGCGAGCGCGCGCGCGAGAGAGCGCGCGCGCGCATATGAGGGTCCG 311
DB 305 CTGCGAGAGAGCGCGAGCGCGCGCGCGAGAGAGCGCGCGCGCATATGAGGGTCCG 364
QY 312 AGCCGAGAAATGACAGCGCAAGAACTCGGCGACGAACTTCAGCTGTGCAAGTGAAG 371
DB 365 AGCCGAGAAATGACAGCGCAAGAACTCGGCGACGAACTTCAGCTGTGCAAGTGAAG 424
QY 372 AGCTGGAAGTGTTCGACACACTCAATACCTGATGTGCCACAGAAGGAACTTG 431
DB 425 AGCTGGAAGTGTTCGACACACTCAATACCTGATGTGCCACAGAAGGAACTTG 484
QY 432 CCGAAACTTAGTGTGATGAAGCAAGTCCGTTGTTAAGAAATGAAGGCGCA 491
DB 485 CCGAAACTTAGTGTGATGAAGCAAGTCCGTTGTTAAGAAATGAAGGCGCA 544
QY 492 GATGTAGCGGACATCAGAGAAATTAATGCTGCGCATGAATAGTGTGCAAGCCAGAG 551
DB 545 GATGTAGCGGACATCAGAGAAATTAATGCTGCGCATGAATAGTGTGCAAGCCAGAG 604
QY 552 ACTGTGTCTACTCATCGTGTGACTAG 577
DB 605 ACTGTGTCTACTCATCGTGTGACTAG 630

RESULT 3

AR339190

LOCUS AR339190

DEFINITION AR339190

ACCESSION AR339190

VERSION AR339190.1 GI:33726047

KEYWORDS

Sequence 681 from patent US 6569662.
PAT 17-AUG-2003

SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 979)
Tang, Y.T., Zhou, P. and Drmanac, R.T.
TITLE Nucleic acids and polypeptides
JOURNAL Patent: US 6569662-A 681 27-May-2003;
FEATURES Location/Qualifiers
source 1..979
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Query Match 98.1%; Score 566; DB 6; Length 979;
Best Local Similarity 100.0%; Pred. No. 4.5e-111;
Matches 566; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

12 GGGCTCCAGCCATGCGCGCTTCGCTCCAGACACCGTGTCTACTGCTGAGTGTAT 71
|
412 GGGCTCCAGCCATGCGCGCTTCGCTCCAGACACCGTGTCTACTGCTGAGTGTAT 471
|
QY 72 ACCAGGTAAATAAGCCCACTCAGCTGGGGGAGCATCAAGCGAAGGCCATG 131
|
Db 472 ACCAGGTAAATAAGCCCACTCAGCTGGGGGAGCATCAAGCGAAGGCCATG 531
|
QY 132 TTGGGCCAAGAGCTCCAGGCTCATGGGTATATGAACCTGAGGGCGGTGGAACCAAG 191
|
Db 532 TTGGGCCAAGAGCTCCAGGCTCATGGGTATATGAACCTGAGGGCGGTGGAACCAAG 591
|
QY 192 AGAAGCGCATGAACCGCATGCGCGCATGATCCCGAGGGCGCGGTGGAACCAAGAGC 251
|
Db 592 AGAAGCGCATGAACCGCATGCGCGCATGATCCCGAGGGCGCGGTGGAACCAAGAGC 651
|
QY 252 CTCGGCAGAGCGCGAGCCCGCGGAGAGCGCGGCCCATGAGGGTCCGC 311
|
Db 652 CTCGGCAGAGCGCGAGCCCGCGGAGAGCGCGGCCCATGAGGGTCCGC 711
|
QY 312 AGCCCGAAGACATGACGAGCGAGCACTCGGCGAGCAAGTTCAAGCTGTGAGGTGAGAG 371
|
Db 712 AGCCCGAAGACATGACGAGCGAGCACTCGGCGAGCAAGTTCAAGCTGTGAGGTGAGAG 771
|
QY 372 AGCTGGAAGATGTTTCCGACACACTCAATACCTGATGTGCCCAAGAGGGAATTG 431
|
Db 772 AGCTGGAAGATGTTTCCGACACACTCAATACCTGATGTGCCCAAGAGGGAATTG 831
|
QY 432 CCGAAGAACTTAGGTGTGACTGAAGACAAAGTGGGTTGTTTAAAGATTAAGAGGCCA 491
|
Db 832 CCGAAGAACTTAGGTGTGACTGAAGACAAAGTGGGTTGTTTAAAGATTAAGAGGCCA 891
|
QY 492 GATGTAGGGGACATCAAGAGAAATTAATGCTGCCCAATGAACCTGCTGACCCGAGAG 551
|
Db 892 GATGTAGGGGACATCAAGAGAAATTAATGCTGCCCAATGAACCTGCTGACCCGAGAG 951
|
QY 552 ACTGTGTACATCGTGTGACTAG 577
|
Db 952 ACTGTGTACATCGTGTGACTAG 977
|

RESULT 4
AX322488/c AX322488 10968 bp DNA linear PAT 07-JAN-2002
LOCUS Sequence 3 from Patent EP162265.
DEFINITION AX322488
ACCESSION AX322488
VERSION AX322488.1 GI:18093554
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
AUTHORS Haendler, B., Weiss, B. and Gesserick, C.
TITLE Human pem as a target for fertility control and Alzheimer therapy
JOURNAL Patent: EP 1162265-A 3 12-DEC-2001;

SOURCE Schering Aktiengesellschaft (DE)
FEATURES Location/Qualifiers
source 1..10968
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 70.9%; Score 409; DB 6; Length 10968;
Best Local Similarity 100.0%; Pred. No. 1.7e-77;
Matches 409; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

12 GGGCTCCAGCCATGCGCGCTTCGCTCCAGACACCGTGTCTACTGCTGAGTGTAT 71
|
Db 8451 GGGCTCCAGCCATGCGCGCTTCGCTCCAGACACCGTGTCTACTGCTGAGTGTAT 8392
|
QY 72 ACCAGGTAAATAAGCCCACTCAGCTGGGGGAGCATCAAGCGAAGGCCATG 131
|
Db 8391 ACCAGGTAAATAAGCCCACTCAGCTGGGGGAGCATCAAGCGAAGGCCATG 8332
|
QY 132 TTGGGCCAAGAGCTCCAGGCTCATGGGTATATGAACCTGAGGGCGGTGGAACCAAG 191
|
Db 8331 TTGGGCCAAGAGCTCCAGGCTCATGGGTATATGAACCTGAGGGCGGTGGAACCAAG 8272
|
QY 192 AGAAGCGCATGAACCGCATGCGCGCATGATCCCGAGGGCGCGGTGGAACCAAGAGC 251
|
Db 8271 AGAAGCGCATGAACCGCATGCGCGCATGATCCCGAGGGCGCGGTGGAACCAAGAGC 8212
|
QY 252 CTCGGCAGAGCGCGAGCCCGCGGAGAGCGCGGCCCATGAGGGTCCGC 311
|
Db 8211 CTCGGCAGAGCGCGAGCCCGCGGAGAGCGCGGCCCATGAGGGTCCGC 8152
|
QY 312 AGCCCGAAGACATGACGAGCGAGCACTCGGCGAGCAAGTTCAAGCTGTGAGGTGAGAG 371
|
Db 8151 AGCCCGAAGACATGACGAGCGAGCACTCGGCGAGCAAGTTCAAGCTGTGAGGTGAGAG 8092
|
QY 372 AGCTGGAAGATGTTTCCGACACACTCAATACCTGATGTGCCCAAG 420
|
Db 8091 AGCTGGAAGATGTTTCCGACACACTCAATACCTGATGTGCCCAAG 8043
|

RESULT 5
AC005023/c 170967 bp DNA linear PRI 21-DEC-1999
LOCUS Home sapiens BAC clone GSI-42113 from Xq25-q26, complete sequence.
DEFINITION AC005023
ACCESSION AC005023
VERSION AC005023.1 GI:3900847
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
AUTHORS Leonard, S., Graves, T. and Coffman, M.
TITLE The sequence of Homo sapiens BAC clone GSI-42113
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 170967)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 170967)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 4 (bases 1 to 170967)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Nov 21, 1998 this sequence version replaced gi:3212968.

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_GS421103

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
This clone was provided for sequencing by Buddy Brownstein in the Center for Genetics in Medicine and by John D. McPherson in the Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
This clone is from the first BAC library from Genome Systems, Inc. (<http://www.genomesystems.com>).
Cell line: lymphoblastoid
Haplotypes: two
VECTOR: pBelobAC
Selection: chloramphenicol
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP3-327A19, 200 bp overlap. The actual start of this clone is at base position 88746 of RP3-327A19; actual end is at 170967 of GS1-42113.
Location/Qualifiers

FEATURES

source
1..170967
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
/map="XG25-q26"
/clone="GS1-42113"
/clone_1fb="GSBAC1"
1..37
/rpt_family="Alu"
38..258
/rpt_family="Alu"
273..568
/rpt_family="Alu"
1789..1835
/rpt_family="polyurpine"
2046..2191
/rpt_family="L1"
2192..2218
/rpt_family="(CAA)n"
2222..2517
/rpt_family="Alu"
2518..2816
/rpt_family="L1"
2817..2889
/rpt_family="Retroviral"
2892..3011
/rpt_family="(TA)n"
3063..3084
/rpt_family="AT-rich"
3085..3493
/rpt_family="Retroviral"
3509..4189
repeat_region

repeat_region /rpt_family="L1"
4190..4481
/rpt_family="Alu"
repeat_region 4482..4662
/rpt_family="L1"
4663..4960
/rpt_family="Alu"
4961..5869
/rpt_family="L1"
5870..5890
/rpt_family="(TAAAA)n"
5891..6624
/rpt_family="L1"
6625..6960
/rpt_family="Retroviral"
6961..7095
/rpt_family="L1"
7096..7274
/rpt_family="Alu"
7275..7509
/rpt_family="L1"
7510..7791
/rpt_family="Alu"
7792..8370
/rpt_family="L1"
8371..8669
/rpt_family="Alu"
8680..9119
/rpt_family="Retroviral"
9120..9194
/rpt_family="MER4-group"
9195..9431
/rpt_family="Retroviral"
9505..9585
/rpt_family="Alu"
9669..9699
/rpt_family="L1"
9701..9721
/rpt_family="(TAA)n"
9722..9996
/rpt_family="Alu"
9997..10338
/rpt_family="L1"
10368..10544
/rpt_family="Alu"
10562..10884
/rpt_family="L1"
10885..11177
/rpt_family="Alu"
11178..11207
/rpt_family="L1"
11208..11512
/rpt_family="Alu"
11513..12024
/rpt_family="L1"
12025..12283
/rpt_family="Alu"
12284..12603
/rpt_family="L1"
12604..12886
/rpt_family="Alu"
12887..12974
/rpt_family="L1"
12975..13118
/rpt_family="Alu"
13121..14063
/rpt_family="L1"
14122..16240
/rpt_family="L1"
16241..16536
/rpt_family="Alu"
16537..19982
/rpt_family="L1"
repeat_region

```

repeat_region 20009..20210
/rep_family="Alu"
repeat_region 20211..20511
/rep_family="Alu"
repeat_region 20512..20617
/rep_family="Alu"
repeat_region 20627..21318
/rep_family="Alu"
repeat_region 21320..21363
/rep_family="AT-rich"
repeat_region 22857..22971
/rep_family="Alu"
repeat_region 23001..23031
/rep_family="GC-rich"
repeat_region 23085..23630
/rep_family="MER4-group"
repeat_region 24345..24373
/rep_family="AT-rich"
repeat_region 24431..24549
/rep_family="Alu"
repeat_region 24762..25060
/rep_family="Alu"
repeat_region 25786..26085
/rep_family="Alu"

```

Query Match 70.9%; Score 409; DB 9; Length 170967;

Best Local Similarity 100.0%; Pred. No. 1.4e-77; Matches 409; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 12 GCGCTCCAGCCATGCGCGCTGCTGTCACAGACACCGTGTCTACTGCTGAGTGTAT 71
DB 168450 GCGCTCCAGCCATGCGCGCTGCTGTCACAGACACCGTGTCTACTGCTGAGTGTAT 168391
QY 72 ACCAGGTAAATAAGAGCCCACTCTGAGTGGGCGACATCAAGCCGAGAGGCCATG 131
DB 168390 ACCAGGTAAATAAGAGCCCACTCTGAGTGGGCGACATCAAGCCGAGAGGCCATG 168331
QY 132 TTGGCCAGAGAGCTCCAGGCTCATGAGGTATATGAACCTGAGGGCGGTGGAACCAAG 191
DB 168330 TTGGCCAGAGAGCTCCAGGCTCATGAGGTATATGAACCTGAGGGCGGTGGAACCAAG 168271
QY 192 AGAAGCGCATGAAACCGCATGAGCGGCGATGATCCCGAGGGCGCGGTGGAACCAAGAGC 251
DB 168270 AGAAGCGCATGAAACCGCATGAGCGGCGATGATCCCGAGGGCGCGGTGGAACCAAGAGC 168211
QY 252 CTCGGCAGCAGCGCCGAGCCCGCGGAGAGCGCGCCCAAGGCGCATGAGAGGTCGCC 311
DB 168210 CTCGGCAGCAGCGCCGAGCCCGCGGAGAGCGCGCCCAAGGCGCATGAGAGGTCGCC 168151
QY 312 AGCCGAGAGACATGACGACGACGAACTCGGCGGACGAGAGTTCAAGCTGTGCAAGTGGAGG 371
DB 168150 AGCCGAGAGACATGACGACGACGAACTCGGCGGACGAGAGTTCAAGCTGTGCAAGTGGAGG 168091
QY 372 AGCTGGAAGTGTTCGACACACTCAATACCTGATGTGCCCAAG 420
DB 168090 AGCTGGAAGTGTTCGACACACTCAATACCTGATGTGCCCAAG 168042

```

RESULT 6
AC145687 139483 bp DNA linear HTG 23-AUG-2003
LOCUS AC145687/c Pan troglodytes chromosome UNK clone CH251-281C1, WORKING DRAFT
DEFINITION
SEQUENCE, 3 unordered pieces.
ACCESSION AC145687
VERSION AC145687.3 GI:34147034
KEYWORDS HTG, HTGS_PHASE1, HTGS_DRAFT, HTGS_ACTIVEFIN.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1 (bases 1 to 139483)
AUTHORS Wilson, R.K.
TITLE The sequence of Pan troglodytes clone

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 139483)
AUTHORS Wilson, R.K.
TITLE Direct Submission
JOURNAL Submitted (19-JUL-2003) Genetics, Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 139483)
AUTHORS Wilson, R.K.
TITLE Direct Submission
JOURNAL Submitted (23-AUG-2003) Genetics, Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT On Aug 23, 2003 this sequence version replaced gi:3387229.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
Project information
Center project name: C_AA0281C01

----- Summary Statistics -----
Sequencing vector: M13, 0%
Sequencing vector: plasmid, 100%
Chemistry: Dye-primer ET, 0% of reads
Chemistry: Dye-terminator Big Dye, 100% of reads
Assembly program: Phrap, version 0.990319
Consensus quality: 137838 bases at least Q40
Consensus quality: 138075 bases at least Q30
Consensus quality: 138228 bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1 1266: contig of 1266 bp in length
* 1267 1366: gap of unknown length
* 1367 51720: contig of 50354 bp in length
* 51721 51820: gap of unknown length
* 51821 139483: contig of 87663 bp in length.

```

FEATURES

```

source
1..139483
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/chromosome="UNK"
/clone="CH251-281C1"
1..1266
/note="assembly_name:Contig5"
1367..51720
/note="assembly_name:Contig17"
51821..139483
/note="assembly_name:Contig18"

```

ORIGIN

Query Match 69.2%; Score 399.4; DB 2; Length 139483;
Best Local Similarity 98.5%; Pred. No. 1.6e-75;
Matches 403; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

QY 12 GCGCTCCAGCCATGCGCGCTGCTGTCACAGACACCGTGTCTACTGCTGAGTGTAT 71
DB 39897 GCGCTCCAGCCATGCGCGCTGCTGTCACAGACACCGTGTCTACTGCTGAGTGTAT 39838
QY 72 ACCAGGTAAATAAGAGCCCACTCTGAGTGGGCGACATCAAGCCGAGAGGCCATG 131
DB 39837 ACCAGGTAAATAAGAGCCCACTCTGAGTGGGCGACATCAAGCCGAGAGGCCATG 39778
QY 132 TTGGCCAGAGAGCTCCAGGCTCATGAGGTATATGAACCTGAGGGCGGTGGAACCAAG 191

```

Db 39777 TTGGCCAGAGAGCTCCAGGCTTCATGAGTAATGAAACCTTGAGGGCGGTGTGAACACAG 39718
 QY 192 AGAAGCGCATGAACCGCGATGCGGCGATGATCCCGAGGGCGGCGGTGAAACACAGAGC 251
 Db 39717 AGAAGCGCATGAACCGCGATGCGGCGATGATCCCGAGGGCGGCGGTGAAACACAGAGC 39658
 QY 252 CTCGCGAGCAGCCGCGAGCCCGCGGAGAGCCGCGCCAGCGGCATGAGAGGTCGCG 311
 Db 39657 CTAGCGAGCAGCCGCGAGCCCGCGGAGAGCCGCGCCAGCGGCATGAGAGGTCGCG 39598
 QY 312 AGCCCGAAGAACATGACGCGACCGAATCTCGGCGACGAAATTCAGCTGTTGACATGAGAG 371
 Db 39597 AGCCCGAAGAACATGACGCGACCGAATCTCGGCGACGAAATTCAGCTGTTGACATGAGAG 39538
 QY 372 AGCTGGAAGATGTTTCCGACACACTCAATACCTGATGTCGCCACAG 420
 Db 39537 AGCTGGAAGATGTTTCCGACACACTCAATACCTGATGTCGCCACAG 39489

RESULT 7
 AC146356/c
 LOCUS 204520 bp DNA linear HTG 27-SEP-2003
 DEFINITION Pongo pygmaeus clone CH253-378A10, WORKING DRAFT SEQUENCE, 6
 ordered pieces.
 AC146356
 AC146356 GI:36430306
 HTG: HTGS PHASE2; HTGS DRAFT.
 KEYWORDS Pongo pygmaeus (orangutan)
 SOURCE Pongo pygmaeus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pongo.

REFERENCE
 AUTHORS 1 (bases 1 to 204520)
 Page, D.C., Rozen, S., Salenz, J.R., Skaletsky, H., Antonellis, A.,
 Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakeley, R.W.,
 Bouffard, G.G., Brinkley, C., Brooks, S., Cariaga, K., Chu, G.,
 Coleman, B., Coleman, H., Engle, J., Granite, S., Guan, X., Gupta, J.,
 Haghighi, P., Han, J., Hansen, N., Ho, S.-L., Hu, P., Hult, B.,
 Idol, J.R., Karlins, E., Kwong, P., Laric, P., Lee-Hin, S.-Q.,
 Legaspi, R., Maduro, Q.L., Maduro, V.B., Marquiles, E.H., Mastello, C.,
 Maskeri, B., McDowell, J., Paguirigan, C., Pearson, R., Portnoy, M.E.,
 Praad, A., Reddik-Dugue, N., Schandler, K., Schueler, M.G., Shah, K.,
 Sison, C., Stancir, S., Thomas, J.W., Thomas, P.J., Tsipouri, V.,
 Vogt, J.L., Wetherby, K.D., Young, A. and Green, E.D.
 NISC Comparative Sequencing Initiative
 TITLE NISC Comparative Sequencing Initiative
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 204520)
 Green, E.D.
 TITLE Direct Submission
 JOURNAL Submitted (13-AUG-2003) NIH Intramural Sequencing Center, 8717
 GroveMont Circle, Gailthersburg, MD 20877, USA
 3 (bases 1 to 204520)
 Green, E.D.
 TITLE Direct Submission
 JOURNAL Submitted (27-SEP-2003) NIH Intramural Sequencing Center, 8717
 GroveMont Circle, Gailthersburg, MD 20877, USA
 On Sep 27, 2003 this sequence version replaced gi:33620791.
 COMMENT
 ----- Genome Center
 Center: NIH Intramural Sequencing Center
 Center code: NISC
 Web site: http://www.nisc.nih.gov
 Contact: nisc_zoo@nhgri.nih.gov
 ----- Project Information
 Center project name: erl
 Center clone name: 378A10

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig, has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average

coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 203385 bases at least Q40
 Consensus quality: 203760 bases at least Q30
 Consensus quality: 203958 bases at least Q20
 Insert size: 211000; agarose-fp
 Insert size: 204020; sum-of-contigs
 Quality coverage: 9.32x in Q20 bases; agarose-fp
 Quality coverage: 9.64x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

* This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

* 1 10013: contig of 10013 bp in length
 * 10014 10113: gap of unknown length
 * 10114 14402: contig of 4289 bp in length
 * 14403 14502: gap of unknown length
 * 14503 79161: gap of 6459 bp in length
 * 79162 79261: gap of unknown length
 * 79262 156374: contig of 77113 bp in length
 * 156375 156474: gap of unknown length
 * 156475 188945: contig of 32471 bp in length
 * 188946 189046: gap of unknown length
 * 189046 204520: contig of 15475 bp in length.

FEATURES

source

1. 204520
 /organism="Pongo pygmaeus"
 /mol_type="genomic DNA"
 /db_xref="taxon:9600"
 /clone="CH253-378A10"
 /clone_1fb="CH253"

misc_feature
 1. 10013
 /note="assembly_fragment"

misc_feature
 clone end:SP6
 vector side:left"

misc_feature
 10114..14402
 /note="assembly_fragment"

misc_feature
 14503..79161
 /note="assembly_fragment"

misc_feature
 79262..156374
 /note="assembly_fragment"

misc_feature
 156475..188945
 /note="assembly_fragment"

misc_feature
 189046..204520
 /note="assembly_fragment"

misc_feature
 missing T7 clone end on 3' end of insert"

ORIGIN

Query Match 66.4%; Score 383.4; DB 2; Length 204520;
 Best Local Similarity 96.1%; Pred. No. 4.2e-72;
 Matches 393; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 12 GCGCTCCAGCAATGGCGCTGCTGCTCAGACACCGTGTCTACCTGAGTGAT 71
 Db 164300 GCGCTCCAGCAATGGCGCTGCTGCTCAGACACCGTGTCTACCTGAGTGAT 164241
 QY 72 ACCAGGTAATAATTAAGCCCAACCTCAGCTGGGCGACATCAAGCGAGGCGCATG 131
 Db 164240 ACCAGGTAATAATTAAGCCCAACCTCAGCTGGGCGACATCAAGCGAGGCGCATG 164181
 QY 132 TTGGCCAGAGAGCTCCAGGCTTCATGAGTAATGAAACCTTGAGGGCGGTGTGAACACAG 191

QY	252	CTGGGGACGACGACCGACACCCCGCGGGAGAGAGCGCGCCAGAGCGGCGCATGAGAGATCCGC	311
Db	76404	CTAGGACAGCTGCGACAGACCC---CTGAGAGAGCCCGGCCAGCGCGGCATGAGAGATCCGC	76466
QY	312	AGCCCCGAGAACTATGCTTTTCCGACACACTCTCAATACCTCTGATGTGGCCCAAG	420
Db	76461	AGCCCGAGAACTATGCTTTTCCGACACACTCTCAATACCTCTGATGTGGCCCAAG	76520
QY	372	AGCTGGAAGTGTCTTTTCCGACACACTCTCAATACCTCTGATGTGGCCCAAG	420
Db	76521	AGCTGGAAGTGTCTTTTCCGACACACTCTCAATACCTCTGATGTGGCCCAAG	76569
RESULT 9			
LOCUS	HS421131	1205 bp	mRNA linear PRI 11-APR-2001
DEFINITION	Novel human mRNA from X chromosome, contains homeobox.		
ACCESSION	AL590524		
VERSION	AL590524.1	GI:13620481	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 1205)		
TITLE	Howell, G.R., Hucklee, E. and Ross, M.T.		
JOURNAL	Direct Submission		
COMMENT	Submitted (09-APR-2001) E-mail contact: humquery@sanger.ac.uk		
	This cDNA sequence was assembled from public domain ESTs and single		

FEATURES	SOURCE
Location/Qualifiers	1. .1205
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/chromosome="X"
	/map="Xq23-Xq24"
	1. .165
exon	/number=1
	87. .953
CDS	/codon_start=1
	/product="hypothetical protein"
	/protein_id="CAC36517.1"
	/db_xref="GI:13620482"
	/db_xref="GOA:Q9BR00"
	/db_xref="SPTREMBL:Q9BR00"
	/translation="MEPPDCSQYMTSLSPAVDKEKELQDMNMTSLTEVEEKEEED
	DAQEPDQGTAAEGKLSKAGQGEEDGAGVPGHLMGNGLEGTSGDQ
	NVEDSDSEKPEQGYRPGANGGLEPQNAQPNVHAFPTIQQLQELCTFQRPQPS
	EFLERRLRASNNVETELAVQIWFENRRKAKRRHQALMARNNLPFMAVGQPMVTAASA
	ITLALFTSGMRDYPFWDHSHSSSLCFPMPPPPSLPLPLMLPMPMPAGQAEFGPP
	FVIAISPTFPMNV"
	166. .577
exon	/number=2
	578. .623
exon	/number=3
	583. .859
misc_feature	/note="matches EST R69179 from clone IMAGE:141589"
	624. .1205
exon	/number=4
	complement(join(730. .956,952. .1203))
misc_feature	/note="matches EST A1954612 from clone IMAGE:2473339"
	complement(739. .1204)
misc_feature	/note="matches EST BB046930 from clone IMAGE:2916991"
	complement(788. .1204)
misc_feature	/note="matches EST A1657125 from clone IMAGE:2244151"
	complement(798. .1205)

misc_feature	/note="matches EST AI209091 from clone IMAGE:1838548"
	complement(811..1200)
	/note="matches EST AA933590 from clone IMAGE:1551106"
ORIGIN	
Query Match	18.5%; Score 106.8; DB 9; Length 1205;
Best Local Similarity	60.7%; Pred. No. 1,1e-12;
Matches 193; Conservative	0; Mismatches 122; Indels 3; Gaps 1;
Dy	203 AACCGCATGAGCGGCATGATCCCGAAGGCGCGCGTGTGAAACCAAGACCTCGGCACGAG 262
Dy	363 AACCTCGAGGCGACCAAGCGGACGAGTGGCAACCTTAGAGCAACGACGACCAAGGAGGAGAG 422
Oy	263 CCGCAGCCCCCGCTGGAGAGAGCCGGCCCGAGCGGCCTATGAGGGTCCGACGCCGAGAC 322
Dy	423 GAACCTTGGAGCAGCATTTTCGCGCCACAGGCGCCCGTCGGGGGGCTGGAACTTCGGCAAC 482
Oy	323 ATGCAAGCACCAACTTCGGCGCACCAAGATTACGCTGTTCAGGTGAGTGGAGAGACTGGAAAGT 382
Dy	483 GCGGAGAGCGCCACAGTCCACGCG---CTTCAACCCCATTTGCAGCTGACAGAGACTGGAGTGC 539
Oy	383 GTTTTCCGACACACTCAATACCTTGATGTGGCCCAAGAAAGGAACTTTGCCGAAAACCTTA 442
Dy	540 ATTTTCCAGCGGACAGTTCCTCCAGTAGTTCCTCGCAAGAGAGACTGGCAAGAGCACTG 599
Oy	443 GGTGTGACTGAGAACAAAGTGGCGGTTTGTTTAAAGATTAAGAGGCGCAGATGTAGCGCA 502
Dy	600 AATGTGACTGAACCTCGCAGTGCAGATTGGTTTGAGATATAGAGCCAAATGTGGAGAGAGA 659
Oy	503 CATCAGAGAGAAATTATG 520
Dy	660 CATCAGAGGCGCATTTATG 677

RESULT	10
LOCUS	AKO58125
DEFINITION	AKO58125. 1194 bp .mRNA linear PRI 12-SEP-2003
ACCESSION	AKO58125
VERSION	AKO58125.1 GI:16554167
KEYWORDS	oligo capping; fis (full insert sequence).
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 Ishibashi,T., Kanehori,K., Yosida,M., Watanabe,S., Ishida,S., Ono,Y., Hottta,T., Hirooka,S., Murakawa,K., Takiguchi,S., Kusano,U., Watanabe,M., Fujimori,K., Tanai,H., Ihida,M., Yamashita,H., Chiba,Y., Suzuki,Y., Hate,H., Nakagawa,K., Mizuno,S., Morinaga,M., Kawamura,M., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Nishikawa,T., Sugiyama,A., Kawakami,B., Nagai,K., Isogai,T. and Sugano.S. NEBO human cdna sequencing project
TITLE	Unpublished
JOURNAL	2 (bases 1 to 1194) Sugano,S. and Suzuki,Y. Direct Submission Submitted (24-OCT-2001) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure, Human Genome Center; Shitokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:flicdn@aims.u-tokyo.ac.jp, Tel :81-3-5449-5286, Fax:81-3-5449-5416) NEBO human cdna sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cdna full insert sequencing: Research Association for Biotechnology (RAB); cdna library construction and 5'-end one pass sequencing: Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure, Human Genome Center; 3'-end one pass sequencing: RAB; clone selection for full insert sequencing: RAB and Helix Research Institute.
AUTHORS	
TITLE	
JOURNAL	
COMMENT	
FEATURES	
SOURCE	/organism="Homo sapiens" 1..1194 Location/Qualifiers

CDS

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="TST02591"
/cisue_type="testis"
/clone_lib="TST"
/notes="Cloning vector: pME18FL3"
84..950
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAB71675.1"
/db_xref="GI:16554168"
/translation="MEPPDQCSQYMTSLSPAVDDKEKLDNNAMVLSLTERVEKEE
DAQPEPBOGTAGEKLSKAGQGEKQGEKQGGAGVPGHLMEDLRETSQSDG
NVEDSDSEKQYRPOGAVGLPEGNQPVNAFTPLQELERIRQREQFPS
EFRRRLRSMNVTLAVQIWFENRRAKRRRQRLMARNMLPFMAVQCPVMTAAE
ITAPLFIQGRDDYFMDHSHSSSLCFMPPPPPSLPLPLMLPPMPAGQAEFGFPF
FVIYPSFTFPNV"

```

ORIGIN

Query Match 18.3%; Score 105.8; DB 9; Length 1194;
 Best Local Similarity 60.6%; Pred. No. 1.8e-12; Mismatches 122; Indels 3; Gaps 1;

```

Matches 192; Conservative 0; Mismatches 122; Indels 3; Gaps 1;

OY 204 ACCGGATGCGCGCATGATCCCGGAGGCGGCGTGAACAACGAGCCTCGGACAGC 263
DB 361 ACCTGAGAGGACACGACGCGCATGCGACGTTGAGGACAGCAGACGAGCGAGAG 420
OY 264 CGCAGCCCCCGCGGAGAGCGCGGCCGCGCATGAGGAGTCCGACGCCGAGAC 323
DB 421 AACCTGGGAGAGCATTCGCGCCACAGGCGCGCTCGGCGGCTGAGCCTGCAACG 480
OY 324 TGCAAGCAGAACTGGGCGCAGAGTTCAGCTGTGAGGTGAGGAGGCTGGAAGTG 383
DB 481 CGCAGACGCCAAGCTCCAGCGC---CTTACCCTCATGAGCTGCGAGGCTGAGCGCA 537
OY 384 TTTTCCGACACATCAATACCTGATGTGCGCCACAAGAGGAACTTGCCGAAACTTAG 443
DB 538 TTTTCAACGCGAGCAGTTCGCCAGTGAAGTCTGCGAAGAGAGGCTGCAAGAGATGA 597
OY 444 GTGTGACGTGAAGACAAAGTGGCGGTTGGTTTAAGAAATAAAGGCCAGATGTAGCGAC 503
DB 598 ATGTGACGTGAACCTCCGACAGTGAATTTGGTTGAGAAATGAAGAGCAATGAGAGAGC 657
OY 504 ATCAGAGAGCAATTAATG 520
DB 658 ATCAGAGGCGCATTAATG 674

```

RESULT 11

LOCUS BC021719 1219 bp mRNA linear PRI 08-OCT-2003
 DEFINITION Homo sapiens PEPP subfamily gene 2, mRNA (cDNA clone MGC:32918
 IMAGE:5269333), complete cds.

ACCESSION BC021719
 VERSION BC021719.2 GI:34194251
 KEYWORDS MGC.
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
 1 (bases 1 to 1219)

Strusberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
 Klausner,R.D., Collins,F.S., Wagner,L., Shemmen,C.M., Schuler,G.D.,
 Altschul,S.F., Zeeberg,B., Buehler,K.H., Schaefer,C.F., Bat,N.K.,
 Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
 Diachenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
 Stapleton,M., Soares,M.B., Bonaldi,M.F., Casavant,T.L.,
 Scheetz,T.E., Brownstein,M.J., Ueda,T.B., Toshiyuki,S.,
 Carinci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
 Abramson,K.J., Mullany,S.J., Bosak,S.A., McEwan,P.J.,
 McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
 Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
 Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,

FAHEY,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S.,
 Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
 Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
 Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
 Buterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalins,D.E.,
 Scherich,A., Schein,J.E., Jones,S.J., and Marra,M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL

22388257
 12477932

REFERENCE

2 (bases 1 to 1219)

AUTHORS

Strusberg,R.

TITLE

Direct Submission

REMARK

Submitted (14-JUN-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

COMMENT

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 On Aug 25, 2003 this sequence version replaced gi:18204436.
 Contact: MGC help desk
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Miklos Palcovits, M.D., Ph.D.
 CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
 Toshiyuki and Piero Carninci (RIKEN)
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
 contact: amadansystemsbio.org
 Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
 Madan, Stephanie Rodriguez, Amy Sanchez and Michelle Whiting

FEATURES

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IBAK Plate: 47 Row: 0 Column: 18
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 14249118.
 Location/Qualifiers

1..1219
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:32918 IMAGE:5269333"
 /tissue_type="testis"
 /clone_lib="NIH_MGC_97"
 /lab_host="DH10B"
 /note="Vector: pBluescript"
 1..1219
 /gene="PEPP-2"
 /db_xref="LOCUSID:84528"
 /db_xref="MIM:300448"
 86..952
 /codon_start=1
 /product="homeobox protein from AU590526"
 /protein_id="AA021719.1"
 /db_xref="GI:18204437"
 /db_xref="LOCUSID:84528"
 /translation="MEPPDQCSQYMTSLSPAVDDKEKLDNNAMVLSLTERVEKEE
 DAQPEPBOGTAGEKLSKAGQGEKQGEKQGGAGVPGHLMEDLRETSQSDG
 NVEDSDSEKQYRPOGAVGLPEGNQPVNAFTPLQELERIRQREQFPS
 EFRRRLRSMNVTLAVQIWFENRRAKRRRQRLMARNMLPFMAVQCPVMTAAE
 ITAPLFIQGRDDYFMDHSHSSSLCFMPPPPPSLPLPLMLPPMPAGQAEFGFPF
 FVIYPSFTFPNV"
 500..658
 /note="homeobox; Region: Homeobox domain"
 /db_xref="CDD:pfam00046"

gene

CDS

misc_feature

ORIGIN

Query Match 18.3%; Score 105.8; DB 9; Length 1219;
 Best Local Similarity 60.6%; Pred. No. 1.8e-12; Mismatches 122; Indels 3; Gaps 1;
 Matches 192; Conservative 0; Mismatches 122; Indels 3; Gaps 1;
 204 ACCGGATGCGCGCATGATCCCGGAGGCGGCGTGAACAACGAGCCTCGGACAGC 263

```

Db      363  ACCCTCAGGGGCAACACCGGAGGATGGCAACCTTAGAGCAAGCGACCAAGAGGAAAGG 422
QY      264  CGCAGCCCCCGCGGAGAGACCGGCCGAGGCGCATGAGAGGATCCGACGCCGAGAAACA 323
Db      423  AACCTGGGAGCAGATTTTCGCCGCCACAGGGGCCCTCGGGGGGCTGGAGCCTGGCAAG 482
QY      324  TGCAGCACAAGACTCGGCGCAGAGATTGCAAGCTGTTCAGAGTGAAGAGCTTGAAAGTG 383
Db      483  CGCAGAGCCCAACGTCACAGC---CTTCACCCCATTTGACGTGACAGAGCTGAGACGCA 539
QY      384  TTTTCGACACACTCAATACCTGATGTGCCCCACAAGAGGAACTTGCAGAAACTTAG 443
Db      540  TTTTCCAGCGACGAGCTTCCCAAGTGAATTCCTGCGAAGAGGCTGGCAGAAAGCATGA 599
QY      444  GTGTGACTGAAGACAAGTGCAGGCTTGGTTTAAGATAAAAGGCGCAGATGAGCGAC 503
Db      600  ATGTGACTGAACCTCGCAGTGCAGATTGGTTTGAAGATTGAAGAGCCAAATGAGAGAGC 659
QY      504  ATCAGAGAGATTATG 520
Db      660  ATCAGAGGCGCATTAATG 676

```

```

RESULT 12
AF317219      1274 bp      mRNA      linear      PRI 13-SEP-2001
LOCUS        Homo sapiens homeobox protein (THG1) mRNA, complete cds.
DEFINITION   AF317219
ACCESSION    AF317219.1 GI:15592921
VERSION      AF317219.1 GI:15592921
KEYWORDS
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

REFERENCE    1 (bases 1 to 1274)
AUTHORS      Yin,Y., Jin,Y. and Levine,A.J.
TITLE        Molecular cloning and characterization of a novel testis homeobox
              gene, THG1, as a potential testicular tumor suppressor gene
              Unpublished
              2 (bases 1 to 1274)
JOURNAL      2 (bases 1 to 1274)
AUTHORS      Yin,Y., Jin,Y. and Levine,A.J.
TITLE        Direct Submission
JOURNAL      Submitted (26-OCT-2000) Radiation Oncology, Columbia University,
              630 West, 168th Street, New York, NY 10032, USA
              Location/Qualifiers
FEATURES     source
              1..1274
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /chromosome="X"
                /tissue="testis"
                1..1274
                /gene="THG1"
                131..997
                /gene="THG1"
                /codon_start=1
                /product="homeobox protein"
                /protein_id="AA102160.1"
                /db_xref="GI:15592922"
                /translation="MEPPDCSQYMTSLSPAVDDEKELQDNNAVTLSTEEYKEEE
                DADPEPEQGTAAAGEKLSKAGAGGEEKDGGGAGVPGHLWEGDLEGTSGSDG
                NWEDSDQSEKEPQOYSRPGAVGLEGEPNAQAPVNAFTPLQLEPERIPQEECPDS
                EFLRRRLASKNVTELAVALQWIFENRRAKWRHRLMARMLPFAVAGQPVNTAAEA
                ITAPLFISGMRDYFWDHSHSSSLCFPMPPPPSLPLMLLPPMPRAGQAFGFPF
                FVIVPSFTFPNV"

```

ORIGIN

```

Query Match      18.3%; Score 105.8; DB 9; Length 1274;
Best Local Similarity 60.6%; Pred. No. 1.8e-12;
Matches 192; Conservative 0; Mismatches 122; Indels 3; Gaps 1;
QY      204  ACCGCATGGCGCATGATCCCGAGGGCGGCGTGGAAACAGAGCCTCGCAGCAGC 263

```

```

Db      408  ACCCTCAGGGGCAACACCGGAGGATGGCAACCTTAGAGCAAGCGACCAAGAGGAAAGG 467
QY      264  CGCAGCCCCCGCGGAGAGACCGGCCGAGGCGCATGAGAGGATCCGACGCCGAGAAACA 323
Db      468  AACCTGGGAGCAGATTTTCGCCGCCACAGGGGCCCTCGGGGGGCTGGAGCCTGGCAAG 527
QY      324  TGCAGCACAAGACTCGGCGCAGAGATTGCAAGCTGTTCAGAGTGAAGAGCTTGAAAGTG 383
Db      528  CGCAGAGCCCAACGTCACAGC---CTTCACCCCATTTGACGTGACAGAGCTGAGACGCA 584
QY      384  TTTTCGACACACTCAATACCTGATGTGCCCCACAAGAGGAACTTGCAGAAACTTAG 443
Db      585  TTTTCCAGCGACGAGCTTCCCAAGTGAATTCCTGCGAAGAGGCTGGCAGAAAGCATGA 644
QY      444  GTGTGACTGAAGACAAGTGCAGGCTTGGTTTAAGATAAAAGGCGCAGATGAGCGAC 503
Db      645  ATGTGACTGAACCTCGCAGTGCAGATTGGTTTGAAGATTGAAGAGCCAAATGAGAGAGC 704
QY      504  ATCAGAGAGATTATG 520
Db      705  ATCAGAGGCGCATTAATG 721

```

```

RESULT 13
HS525N141      1297 bp      mRNA      linear      PRI 11-APR-2001
LOCUS        Novel human mRNA from X chromosome, contains homeobox.
DEFINITION   HS525N141
ACCESSION    AL590526
VERSION      AL590526.1 GI:13620840
KEYWORDS
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

REFERENCE    1 (bases 1 to 1297)
AUTHORS      Howell,G.R., Huckle,E. and Rose,M.T.
TITLE        Direct Submission
JOURNAL      Submitted (09-APR-2001) E-mail contact: humquery@sanger.ac.uk
              This cDNA sequence was assembled from public domain ESTs and single
              pass sequencing reads from expressed DNA templates, aligned to the
              genomic DNA sequence from the bacterial clone 525N14 (AC002086).
              The EST sequences listed match this sequence with an identity of at
              least 95% between the coordinates shown.
              Further information can be found at
              http://www.sanger.ac.uk/HGP/Chrx/ Sanger Centre name :
              dJ525N14.CX.1
              Location/Qualifiers
FEATURES     source
              1..1297
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /chromosome="X"
                /map="Xq23-Xq24"
                1..269
                /number=1
                191..1057
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="CAC36519.1"
                /db_xref="GI:13620841"
                /db_xref="GOA:Q9BOY4"
                /db_xref="SPTREMBL:Q9BOY4"
                /translation="MEPPDCSQYMTSLSPAVDDEKELQDNNAVTLSTEEYKEEE
                DADPEPEQGTAAAGEKLSKAGAGGEEKDGGGAGVPGHLWEGDLEGTSGSDG
                NWEDSDQSEKEPQOYSRPGAVGLEGEPNAQAPVNAFTPLQLEPERIPQEECPDS
                EFLRRRLASKNVTELAVALQWIFENRRAKWRHRLMARMLPFAVAGQPVNTAAEA
                ITAPLFISGMRDYFWDHSHSSSLCFPMPPPPSLPLMLLPPMPRAGQAFGFPF
                FVIVPSFTFPNV"

```

```

exon      270..681
exon      682..727
exon      /number=2
exon      /number=3

```

```
misc_feature      687..963
                  /note="matches EST R69179 from clone IMAGE:141589"
exon              728..1297
                  /number=4
misc_feature      /note="complement(join(834..1060,1056..1297))"
misc_feature      /note="matches EST A1954612 from clone IMAGE:247139"
misc_feature      /note="complement(843..1297)"
misc_feature      /note="matches EST BE046930 from clone IMAGE:2916991"
misc_feature      /note="complement(892..1297)"
misc_feature      /note="matches EST A1657125 from clone IMAGE:2244151"
misc_feature      /note="complement(902..1297)"
misc_feature      /note="matches EST A1209091 from clone IMAGE:1838548"
misc_feature      /note="complement(915..1297)"
misc_feature      /note="matches EST A933590 from clone IMAGE:1551106"
misc_feature      /note="complement(1066..1296)"
                  /note="matches EST R69071 from clone IMAGE:141589"

ORIGIN
Query Match      18.3%; Score 105.8; DB 9; Length 1297;
Best Local Similarity 60.6%; Pred. No. 1.8e-12;
Matches 192; Conservative 0; Mismatches 122; Indels 3; Gaps 1;

QY 204 ACCGCGATGCGCGCATGATCCCGAGCGCGCGGTGAAACCAAGAGCCTCGGACAGC 263
DB 468 ACCTCGAGGGCACACGCGGACGATGGCAACGTTAGAGCACAGCAGACGAGAGAGG 527
QY 264 CCCAGCCCCCGCGAGAGCCGCCAGCCGCTGAGAGGTCCGCGACCCGAGAAACA 323
DB 528 AACCTGGGAGAGATGTTGGCGCCGACAGGCGCCCTCGGGGGCTGGAGCCTGGAGAG 587
QY 324 TCGAGCCAGAACTCGGCGCACAGAACTGACGCTGTTGAGAGTGAGAGAGCTGAAATG 383
DB 588 CGCAGAGAGCCCAACGTCACACG---CTTACCCCATTTGACGTCGAGAGCTGAGAGCGCA 644
QY 384 TTTTCGACACACTCAATACCTGATGTGCCCAAGAAAGAACTTGCCGAAACTTGA 443
DB 645 TTTTCCACGCGACGAGCTGCCAGATGAGTCTCTGGAAGAGCTGGCAAGAGCATGA 704
QY 444 GGTGATCTAAGACAAAGTGGCGTTTGTGTTAAAGTAAGGCGCAATGATAGCCGAC 503
DB 705 ATGTGACTGAACCTCGAGTGCAGATTGTTGATGAATGAAAGCAATATGAGAGAGAC 764
QY 504 ATCAGAGAGATTATATG 520
DB 765 ATCAGAGGCAATTATATG 781

RESULT 14
AY114148          1451 bp      mRNA      linear      PRI 29-JUN-2002
LOCUS             Homo sapiens ESX1L (ESX1L) mRNA, complete cde.
DEFINITION        AY114148
ACCESSION         AY114148.1 GI:21629645
VERSION           AY114148.1 GI:21629645
KEYWORDS
SOURCE            Homo sapiens (human)
ORGANISM          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE         1 (bases 1 to 1451)
AUTHORS          Pohn,L.E. and Behringer,R.R.
TITLE             ESX1L, a novel X chromosome-linked human homeobox gene expressed in
                  the placenta and testis
JOURNAL           Genomics 74 (1), 105-108 (2001)
MEDLINE           21269184
PUBMED           11374906
REFERENCE         2 (bases 1 to 1451)
AUTHORS          Pohn,L.E. and Behringer,R.R.
TITLE             Direct Submission
JOURNAL           Submitted (17-MAY-2002) Molecular Genetics, MD Anderson Cancer
                  Center, 1515 Holcombe-Box 45, Houston, TX 77030-4095, USA
FEATURES          1..1451
                  Location/Qualifiers
                  /organism="Homo sapiens"
```

```
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="X"
/map="Xq22.1-q22.3"
/tissue_type="pooled germ cell tumors"
/note="genomic sequence deposited in GenBank Accession Number AL049631; derived from dbEST Accession Number A1964078"
1..1451
/gene="ESX1L"
44..1264
/gene="ESX1L"
/note="ESX1L; expressed in placenta and testis"
/codon_start=1
/product="ESX1L"
/protein_id="AA062141.1"
/db_xref="GI:21629646"
/translation="MESLRGYTHSDIGYRSLAVGSDIEVDNDEKLTIVSLMARGEDB
ENTRSKPEYGTAEENNVTGEGSVDDDDREGGHEPQQOEEPLTYEQOEEPP
LLELKQOBERPPQTVEGPQPAEGPOTAEQPPERKRRRTAFQFOLELENPDE
SOYPRVVARERHAAHLNLTEDRVQVFRKAKKRNQRLMLRPTATDLAHPDMF
LGGATYAAPALDPALCVHLVPLRPVLVPPMPRRPMPVMPRPPIAPMPMAVP
PPGSMAPVPPGPMAVPPMPMAVPPMPMAVPPMPMAVPPGPMAVPPGPMAVPPGP
MAPVINSYVACPFP"
458..640
/gene="ESX1L"
/note="Region: homeodomain"
1432..1437
/gene="ESX1L"

ORIGIN
Query Match      17.9%; Score 103; DB 9; Length 1451;
Best Local Similarity 59.3%; Pred. No. 6.9e-12;
Matches 194; Conservative 0; Mismatches 130; Indels 3; Gaps 1;

QY 246 AGAAGCTCTGGAGACAGCCGACCCCGGAGAGGCGGCCAGCGGCATGAGAG 305
DB 378 AGCCGCCCGACAGACGCGTGAAGGGGCCACAGCCGCGGAGGCGCAAAACCGCTGAG 437
QY 306 GTCCGAGGCCGAGAACATGACCAAGCACTCGGCGCAGAACTTACGCTGTGGCAG 365
DB 438 GACCAAGGCCCGCAGAGAGAA---ACGCGCGCGCGCACCGCTTACAGCATTTTCA 494
QY 366 TGAAGAGCTGGAAGATGTTTCCGACACATCAATACCTGATGTGCCACAAAGAG 425
DB 495 TCAGAGAGCTAGAGAACTTTTCGATGATCTCAATATCCGACGTTGTGGCGCGAGAA 554
QY 426 AACTTGCCGAAACTTAGGTGACTGAAGACAAAGTGGGGTTGTTAAGATTAATA 485
DB 555 GACTTGCAGACAGCTGAAATTTGACTGAAGACAGATGAGATTTGTTTCAAGACAGAA 614
QY 486 GGGCCAGATGTAAGGCGACATCAGAGAAATTAATGCTGCCAATGAATACGTGCTGACC 545
DB 615 GAGCCAAAGTGAAGAAAGAAATCAGAGGGGTCTATGTTGAGAAACACTGCTACTGCTGACC 674
QY 546 CAGAGCATCTGTCTACATCTGCTGAG 572
DB 675 TGCCCAACCTTTGACATGTTCTTGG 701

RESULT 15
AX834967          1518 bp      DNA      linear      PAT 15-DEC-2003
LOCUS             Sequence 2091 from Patent EP1347046.
DEFINITION        AX834967
ACCESSION         AX834967
VERSION           AX834967.1 GI:39921102
KEYWORDS
SOURCE            Homo sapiens (human)
ORGANISM          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE         1
```

AUTHORS

Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S., Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R., Tamachi, I., Seki, N., Yoshikawa, T., Otsuka, M., Negahari, K. and Masuko, Y.

TITLE

Full-length cDNA sequences

JOURNAL

Patent: EP 1347046-A 2091 24-SEP-2003;

FEATURES

Research Association for Biotechnology (JP)

source

1. 1518
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 17.9%; Score 103; DB 6; Length 1518;

Best Local Similarity 59.3%; Pred. No. 6, 9e-12;

Matches 194; Conservative 0; Mismatches 130; Indels 3; Gaps 1;

```
QY 246 AGGAGCCTCGGCGACGAGCCGCGAGCCCGCGGAGAGAGCCGCGCCAGCGGCCCATGAGG 305
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 419 AGCGCGCCCGACGACCGGTGAGGGGCCACAGCCGCGCGGAGGGGCCCAAAACCGCTGAGG 478
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 306 GTCCGAGGCCCGAGAACATGACCCAGCACTGGGCGCAGAACTTCAGCTTTGACG 365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 479 GACCAAGCCCGCAGAGAGGAA--ACGCGCGCGCGCACCGGTCACGCACTTCAGC 535
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 366 TGGAGAGCTGGAAGTGTTCGACACACTCAATATCCTGATGTGCCACAGAAGAG 425
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 536 TGCAGAGAGCTAGGAAGCTTTTCGATGAATCTCAATATCCGACGTTGTGGCGCGAGAG 595
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 426 AACTTGCCGAAAACCTTAGGTGACTGAGAGCAAGTGGCGGTTTGTTTAAAGATPAAA 485
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 596 GACTTGACACACCGCTGAATTGTACTGAAAGACAGAGTGACAGTTTGTTTCAGAACAGAA 655
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 486 GGGCCGAGTGTGGCGACATCAGAGAGAAATATGCTCGCCATGAACTAAGTGCAGACC 545
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 656 GAGCCAAAGTGAAACGAAATCAGAGGGTGTATGTTGAGAAACACTGCTACTGCTGACC 715
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 546 CAGACGACTGTGTCTACATCGTGTGG 572
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 716 TGGCCCAACCTTTGACATGTTCTTGG 742
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: June 18, 2004, 04:38:10
Job time : 2488 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 16, 2004, 19:00:11 ; Search time 44.5 Seconds

(without alignments)
7305.886 Million cell updates/sec

Title: US-09-867-753-1

Perfect score: 1050
Sequence: 1 tccacatcagcgcgtccag.....ctcacatcgtcgtgactag 577

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1158786 seqs, 281726120 residues

Total number of hits satisfying chosen parameters: 2317572

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-Q=/cgn2_1/USPTO_epool_P/US09867753/runac.16062004.163637.27855/app_query.fasta.1.775
-DB=Published Applications AA -OPMT=fastan -SUFFIX=xapb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bite -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MOD=LOCAL -OUTPMT=ptc -SCORE=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09867753@cgn2_1.1.13@runac.16062004.163637.27855
-NCPU=6 -ICPU=3 -NO MMAP -LARGEOQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA.*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pcp.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pcp.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pcp.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pcp.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pcp.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pcp.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pcp.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pcp.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pcp.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pcp.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pcp.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pcp.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pcp.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pcp.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pcp.*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pcp.*
17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pcp.*
18: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pcp.*
19: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	1023	97.4	194	9	US-09-764-864-1155	Sequence 1155, Ap
2	994	94.7	194	9	US-09-764-864-1568	Sequence 1568, Ap
3	986	93.9	184	9	US-09-867-753-2	Sequence 2, Appli
4	298.5	28.4	415	15	US-10-108-260A-4534	Sequence 4534, Ap
5	196	18.7	562	15	US-10-295-027-676	Sequence 676, App
6	196	18.7	562	15	US-10-173-999-133	Sequence 133, App
7	190	18.1	1506	14	US-10-161-051-88	Sequence 88, Appl
8	187.5	17.9	215	12	US-10-221-625-88	Sequence 88, Appl
9	184.5	17.6	217	14	US-10-301-822-165	Sequence 165, App
10	184.5	17.6	245	14	US-10-301-822-167	Sequence 167, App
11	178	17.0	190	14	US-10-158-160A-16	Sequence 16, Appl
12	176.5	16.8	225	14	US-10-158-160A-13	Sequence 13, Appl
13	176.5	16.8	292	14	US-10-158-160A-11	Sequence 11, Appl
14	176.5	16.8	489	12	US-10-087-192-1983	Sequence 1983, Ap
15	176	16.8	479	14	US-10-245-171A-7	Sequence 7, Appli
16	176	16.8	479	14	US-10-245-171A-11	Sequence 11, Appl
17	174	16.6	234	10	US-09-105-470-4	Sequence 4, Appli
18	174	16.6	234	15	US-10-614-275-4	Sequence 4, Appli
19	172	16.4	467	14	US-10-245-171A-5	Sequence 5, Appli
20	172	16.4	504	12	US-10-087-192-1986	Sequence 1986, Ap
21	172	16.4	520	14	US-10-245-171A-2	Sequence 2, Appli
22	172	16.4	520	14	US-10-245-171A-1	Sequence 1, Appli
23	172	16.4	520	14	US-10-245-171A-3	Sequence 3, Appli
24	172	16.4	520	14	US-10-245-171A-4	Sequence 4, Appli
25	172	16.4	524	14	US-10-245-171A-6	Sequence 6, Appli
26	171.5	16.3	289	14	US-10-097-340-71	Sequence 71, Appl
27	171.5	16.3	289	15	US-10-295-027-626	Sequence 626, App
28	170.5	16.2	70	8	US-08-701-278-10	Sequence 10, Appl
29	169.5	16.1	332	14	US-10-245-171A-10	Sequence 10, Appl
30	169	16.1	241	9	US-09-759-143-842	Sequence 842, App
31	169	16.1	241	9	US-09-780-669-842	Sequence 842, App
32	169	16.1	241	9	US-09-822-827-842	Sequence 842, App
33	169	16.1	241	9	US-09-895-793-842	Sequence 842, App
34	169	16.1	241	9	US-09-895-814-842	Sequence 842, App
35	169	16.1	241	13	US-10-012-896-842	Sequence 842, App
36	169	16.1	241	14	US-10-144-678A-842	Sequence 842, App
37	169	16.1	241	14	US-10-294-025-842	Sequence 1, Appli
38	167	15.9	60	14	US-10-158-160A-1	Sequence 1, Appli
39	167	15.9	60	14	US-10-105-004-18	Sequence 18, Appl
40	167	15.9	60	14	US-10-105-004-19	Sequence 19, Appl
41	166	15.8	70	8	US-08-701-278-4	Sequence 4, Appli
42	166	15.8	309	12	US-10-408-501-6	Sequence 6, Appli
43	166	15.8	314	12	US-10-408-501-2	Sequence 2, Appli
44	166	15.8	373	16	US-10-408-765A-975	Sequence 975, App
45	165.5	15.8	263	8	US-08-701-278-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-764-864-1155
Sequence 1155, Application US/09764864
Patent No. US20020132753A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT23
CURRENT APPLICATION NUMBER: US/09/764,864
PRIORITY FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 1792
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 1155
LENGTH: 194
TYPE: PRT
ORGANISM: Homo sapiens
US-09-764-864-1155

Alignment Scores:
Pred. No.: 1.38e-73
Score: 1023.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%

Length: 194
Matches: 191
Conservative: 0
Mismatches: 0

Query Match: 97.43% Indels: 0
DB: 9 Gaps: 0
US-09-867-753-1 (1-577) x US-09-764-864-1155 (1-194)
QY 2 CCAACATCAGGCGCTCCAGCATGCGCGCTGCTGCTCCAGACACCGTGTCTACTGC 61
DB 4 ProthSergLYAlaProAlaMetAlaArgSerLeuValHisAspThrValPheTyrCys 23
QY 62 CTGAGGTATACAGGTTAAATAATAGCCCACTCAGCTGGGGCAGCATCAACGCA 121
DB 24 LeuSerValTyrGlnValIlysIleSerProThrProGlnLeuGlyAlaIaSerSerAla 43
QY 122 GAAGGCATGTTGGCCAAAGAGCTCCAGGCTCATGGTAATATGAACCTGAGGGCGGT 181
DB 44 GlnGlyHisValGlyGlnGlyAlaProGlyLeuMetGlyAsnMetAsnProGlnGlyGly 63
QY 182 GTGAACACAGAGAACCGCATGAAACCGCATGCGCGCATGATCCCGAGGGCGCGGTGA 241
DB 64 ValAsnHisGlyAsnGlyMetAsnArgAspGlyGlyMetLeProGlnGlyGlyGly 83
QY 242 AACCAAGAGCTCGGAGAGCGCGAGCCCGCGCGAGAGAGGGCGCGCGCATG 301
DB 84 AsnGlnGlnProArgGlnGlnProGlnProProProGlnGlnProAlaGlnAlaMet 103
QY 302 GAGGTCGCGAGCCCGAGAACATGACGCCAGAACTCGGCGAGCAAGATTACGCTGTTG 361
DB 104 GlnGlyProGlnProGlnProGlnMetGlnProArgThrArgThrIlySphThrLeu 123
QY 362 CAGGTGGAGAGCTGGAAGTGTCTTCCGACACATCATACCTGTATGTCGCCACAAAG 421
DB 124 GlnValGlnGlnLeuGlnSerValPheArgHisThrGlnIlyProAspValProThrArg 143
QY 422 AGGGAACCTGCGCAAACTTAGGTGACTGACAGAGAAATTAATGCTGCCAATGAATACGTGCT 481
DB 144 ArgGlnLeuAlaGlnAsnLeuGlyValThrGlnAspIlyValAlaTyrPheIlyAsn 163
QY 482 AAAAGGCCAGATGTAGGCGCATGACAGAGAAATTAATGCTGCCAATGAATACGTGCT 541
DB 164 LysArgAlaArgCysArgArgHisGlnArgGlnLeuMetLeuAlaAsnGlnLeuArgAla 183
QY 542 GACCCAGAGCACTGTGTCTTACATCGTCGTGGAC 574
DB 184 AspProAspAspCysValTyrIleValValAsp 194
RESULT 2
US-09-764-864-11568
; Sequence 1568, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1568
; LENGTH: 194
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (16)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (21)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (22)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE

LOCATION: (25)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (157)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-11568
Alignment Scores:
Pred. No.: 2,9e-71 Length: 194
Score: 994.00 Matches: 186
Percent Similarity: 97.38% Conservative: 0
Best Local Similarity: 97.38% Mismatches: 5
Query Match: 94.67% Indels: 0
DB: 9 Gaps: 0
US-09-867-753-1 (1-577) x US-09-764-864-11568 (1-194)
QY 2 CCAACATCAGGCGCTCCAGCATGCGCGCTGCTGCTCCAGACACCGTGTCTACTGC 61
DB 4 ProthSergLYAlaProAlaMetAlaArgSerLeu**HisAspThrVal*****Cys 23
QY 62 CTGAGGTATACAGGTTAAATAATAGCCCACTCAGCTGAGGGCAGCATCAACGCA 121
DB 24 Leu***ValTyrGlnValIlysIleSerProThrProGlnLeuGlyAlaIaSerSerAla 43
QY 122 GAAGGCATGTTGGCCAAAGAGCTCCAGGCTCATGGTAATATGAACCTGAGGGCGGT 181
DB 44 GlnGlyHisValGlyGlnGlyAlaProGlyLeuMetGlyAsnMetAsnProGlnGlyGly 63
QY 182 GTGAACACAGAGAACCGCATGAAACCGCATGCGCGCATGATCCCGAGGGCGCGGTGA 241
DB 64 ValAsnHisGlyAsnGlyMetAsnArgAspGlyGlyMetLeProGlnGlyGlyGly 83
QY 242 AACCAAGAGCTCGGAGAGCGCGAGCCCGCGCGAGAGAGGGCGCGCGCATG 301
DB 84 AsnGlnGlnProArgGlnGlnProGlnProProProGlnGlnProAlaGlnAlaMet 103
QY 302 GAGGTCGCGAGCCCGAGAACATGACGCCAGAACTCGGCGAGCAAGATTACGCTGTTG 361
DB 104 GlnGlyProGlnProGlnProGlnMetGlnProArgThrArgThrIlySphThrLeu 123
QY 362 CAGGTGGAGAGCTGGAAGTGTCTTCCGACACATCATACCTGTATGTCGCCACAAAG 421
DB 124 GlnValGlnGlnLeuGlnSerValPheArgHisThrGlnIlyProAspValProThrArg 143
QY 422 AGGGAACCTGCGCAAACTTAGGTGACTGACAGAGAAATTAATGCTGCCAATGAATACGTGCT 481
DB 144 ArgGlnLeuAlaGlnAsnLeuGlyValThrGlnAspIlyValAlaTyrPheIlyAsn 163
QY 482 AAAAGGCCAGATGTAGGCGCATGACAGAGAAATTAATGCTGCCAATGAATACGTGCT 541
DB 164 LysArgAlaArgCysArgArgHisGlnArgGlnLeuMetLeuAlaAsnGlnLeuArgAla 183
QY 542 GACCCAGAGCACTGTGTCTTACATCGTCGTGGAC 574
DB 184 AspProAspAspCysValTyrIleValValAsp 194
RESULT 3
US-09-867-753-2
; Sequence 2, Application US/09867753
; Patent No. US20020046764A1
; GENERAL INFORMATION:
; APPLICANT: WEISS, BERTRAM
; APPLICANT: GESSERICK, CHRISTOPH
; APPLICANT: HAENDLER, BERNARD
; TITLE OF INVENTION: HUMAN PEM AS A TARGET FOR BIRTH CONTROL AND TREATMENT
; FILE REFERENCE: SCH-1810
; CURRENT APPLICATION NUMBER: US/09/867,753
; CURRENT FILING DATE: 2001-05-31
; Prior application data removed - consult PALM or file wrapper
; PRIOR FILING DATE: 2000-05-31
; NUMBER OF SEQ ID NOS: 6

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 184
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-867-753-2

Alignment Scores:
  Pred. No.: 1,26e-70      Length: 184
  Score: 986.00           Matches: 184
  Percent Similarity: 100.00%  Conservative: 0
  Best Local Similarity: 100.00%  Mismatches: 0
  Query Match: 93.90%         Indels: 0
  Gaps: 0

US-09-867-753-1 (1-577) x US-09-867-753-2 (1-184)

QY 23 ATGGCGCGTTCCTGCTCCACGACCGTGTCTACTGCTGATGATATACAGGTAAGA 82
DB 1 MetAlaArgSerLeuValHisAspThrValPheTyrCysLeuSerValTyrGlnValLys 20
QY 83 ATAAAGCCCAACCTTCAGCTGGGCGGAGATCAAGCGCAAGGCGCATTTGGCCAGA 142
DB 21 TieserProthrProGlnLeuGlyAlaAlaSerSerAlaGlnGlyHisValGlnGlnGly 40
QY 143 GCTCCAGGCGCTCATGGGTATATGAACCTTGAGGGCGGTGTGAACCAAGAGACGGCATG 202
DB 41 AlaProGlyLeuMetCysIleMetLeuAsnProGlnGlyGlyValHisHisGlnGlnGlnGln 60
QY 203 AACCGGATGGCGGCGCATGTATCCCGAGGGCGGCGGTGAACCAAGAGACCTTCGCGCAG 262
DB 61 AsnAlaGArgGlyGlyMetIleProGlnGlyGlyGlyValAsnGlnGlnGlnGlnGlnGln 80
QY 263 CCGGAGCGCCCGCGGAGGCGCGGCGGCGGCGCATGAGAGGTCCGAGCGCCGAGAAC 322
DB 81 ProGlnProProProGlnGlnProAlaGlnAlaAlaMetCysGlnGlyProGlnProGlnAsn 100
QY 323 ATGACGACCAAGAACTCGGCGCAAGATTCAAGCTTTGACAGGTGAGAGAGCTGGAAGT 382
DB 101 MetGlnProArgThrArgThrArgThrArgThrArgThrArgThrArgThrArgThrArgThr 120
QY 383 GTTTTCCGACACATCAATATACCTGATGTGTGCCACAAAGAGGAACTTCCGAGAACTTA 442
DB 121 ValIlePheArgHisThrGlnTyrProAspValProThrArgArgGlnLeuAlaGlnLeu 140
QY 443 GGTGTGACCTGAGACAAAGTGGGGGTTTGGTTTAAGATTAAGAGGCGCAGATGAGCGCA 502
DB 141 GlyValThrGlnAspLysValAlaGlyAlaTyrPheLysAsnLysArgAlaArgCysArgArg 160
QY 503 CATCAGAGAGATTAATGCTCGCAATGAACCTACGCTGACCCGAGAGAGCTGTGTCTAC 562
DB 161 HisGlnArgGlnLeuMetLeuAlaAsnGlnLeuArgAlaAspProAspAspCysValTyr 180
QY 563 ATCGTCGTGGAC 574
DB 181 IleValValAsp 184

RESULT 4
US-10-108-260A-4534
; Sequence 4534, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1el full length cDNA
; FILE REFERENCE: HI-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4534
; LENGTH: 415
; TYPE: PRF
; ORGANISM: Homo sapiens
```

```
US-10-108-260A-4534

Alignment Scores:
  Pred. No.: 1,46e-15      Length: 415
  Score: 298.50           Matches: 71
  Percent Similarity: 49.21%  Conservative: 22
  Best Local Similarity: 37.57%  Mismatches: 39
  Query Match: 28.43%         Indels: 57
  Gaps: 6

US-09-867-753-1 (1-577) x US-10-108-260A-4534 (1-415)

QY 170 CCTGAGGCGGCGGTGTGAACCAAGAGACCGGATGAGCGGCGCATGATCCCG--- 226
DB 51 ProGlnTyrGlyThrGlnAlaGlnLeuAsnValGlnThrGlnGlySer---ValProSer 69
QY 227 -----GAGGCGCGCGGTGGAACCAAGAGACCTTCGCGAGCAGCCGACGCC 271
DB 70 AspAspGlnAspArgGlnGlyGlyGlyGlyHisGlnProGlnGlnGlnGlnGlnGlnPro 89
QY 272 CCG----- 274
DB 90 ProLeuThrLysProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 109
QY 275 CCGGAGAGCGCGCGCCGCGCGCATGAGAGGTCCGACGCC----- 316
DB 110 GlnGlnGlnProProGlnThrThrValGlnGlnProGlnProGlnGlnGlnGlnGlnGln 129
QY 317 ---GAGAACATGCACCA-----CGAACTCGGCGCAGAAAGTTCAGCTGTG 361
DB 130 AlaGlnGlyProGlnProProGlnLysArgArgArgArgThrAlaPheThrGlnPhe 149
QY 362 GAGGAGGAGGAGCTGGAAGAGTGTTCGACACATCAATACCTGATGTGCCACAGA 421
DB 150 GlnLeuGlnGlnLeuGlnAsnPhePheAspGlnSerGlnTyrProAspValAlaLys 169
QY 422 AGGGAACCTTGCCGAACCTTGAAGTGTGAAGAGCAAGAGTGGCGGTGTTTAAGAT 481
DB 170 GlnArgLeuAlaAlaArgLeuAsnLeuThrGlnAspArgValGlnValTyrPheGlnAsn 189
QY 482 AAAAGGCCAGATGTAGCGCGACATCAGAGAGATTAATGCTGCCAATGAACCTGTCT 541
DB 190 ArgArgAlaLysTyrLysArgAsnGlnArgValIleuMetLeuArgAsnThrAla 209
QY 541 ----- 541
DB 210 AspLeuAlaHisProLeuAspMetPheLeuGlnGlyAlaTyrTyrAlaAlaProAlaLeu 229
QY 542 GACCCAGACGACTGTGTCTACATCGTC 568
DB 230 AspProAlaLeuCysValHisLeuVal 238

RESULT 5
US-10-295-027-676
; Sequence 676, Application US/10295027
; Publication No. US2003023250A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glyne, Richard
; APPLICANT: Hevez, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Susan R.
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
```



```

QY 224 CCGAGGCGCGGTGGAAC----- 244
DB 215 ProAlaAlaGlyGlyThrGlyThrGlyUaAspArgGluGluLeuLeuGluUaAspGlu 234
QY 244 ----- 244
DB 235 GluAspGluAspGluGluGluLeuLeuGluUaAspArgGluGluLeuLeuGluUaAsp 254
QY 245 -----CAGAGCGCTGGGAGAGAGCGG----- 265
DB 255 AspAlaArgAlaLeuLeuUaAspGluProAlaArgCysProValAlaAlaThrGlyAlaVal 274
QY 266 -----CAGCGCGCGCGCGAGAG 283
DB 275 AlaAlaAlaAlaAlaAlaAlaAlaThrGluGlyGlyGluLeuSerProGlyGluGlu 294
QY 284 -----CCGCGCCAGGCGCGCGAGAGGTCGCCAGCGCGAGAAC----- 322
DB 295 LeuLeuLeuHisProGluUaAspAla-----GluGlyUaAspGlyGluUaAspSerValCys 312
QY 323 -----ATGAGCGCAGAGACTCGCGCG 343
DB 313 LeuSerAlaGlySerAspSerGluGluGlyLeuLeuUaAspArgGlyGlyGlyArg 332
QY 344 AGAAGTTTCAGCGCTGTTCAGAGTGAGAGAGCTGAAAGTGTTCGAGACACTCAATAC 403
DB 333 ThrThrPheThrSerGlyGlyGluGluGluUaArgAlaPheGlnHisThrHisGly 352
QY 404 CCTGATGTCGCCAGAGAGGAACTTGGCGAAACTTGGGTGCTGACGTGAGCAAAAGTG 463
DB 353 ProAspValPheThrArgGluGluGluUaMetArgLeuAspLeuThrGluAlaArgVal 372

QY 464 CGGGTTGGTTTAAAGATTAAGCGCCAGATGATGAGCGCAGATCAGAGAGA 511
DB 373 GlnValTTPPheGlnMetArgAlaGlyTTPArgGlyArgGluUa 388

RESULT 7
US-10-161-051-88
; Sequence 88, Application US/10161051
; Publication No. US20030152945A1
; GENERAL INFORMATION:
; APPLICANT: Peter Deak
; APPLICANT: David M Glover
; APPLICANT: Carol Mdgley
; TITLE OF INVENTION: Cell cycle progression proteins
; FILE REFERENCE: CCI-021CP
; CURRENT APPLICATION NUMBER: US/10/161,051
; CURRENT FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: GB 0007268.6
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 88
; LENGTH: 1506
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-161-051-88

Alignment Scores:
Pred. No.: 7.88e-07 Length: 1506
Score: 190.00 Matches: 55
Percent Similarity: 43.86% Conservative: 20
Best Local Similarity: 33.16% Mismatches: 60
Query Match: 18.10% Indels: 36
DB: 14 Gaps: 7

US-09-867-753-1 (1-577) x US-10-161-051-88 (1-1506)
QY 92 AACCTCAGCTGGGGGACATCAACGAGAGGAGCATTTGGCCAGAGAGCTCCAGGC 151
DB 1123 ThrProProGlyGlyGlyProAlaGlyAlaGlyAlaLeuGlnProGlyGlySerGly 1142
QY 152 CTCATG-----GATAATATG-----AACCT-----GAGGCGGTGTG 184

```

```

DB 1143 SerSerTyGlySerAspGlyAsnMetSerSerAsnProAsnSerSerAsnThr 1162
QY 185 AACACAGAGAACGGCATGACCGCATGGCGG----- 217
DB 1163 ThrHisSerAsnGlyHisAsnThrAsnSerGlySerGlyCysGlyAspSerAlaGly 1182
QY 218 -----ATGATCCCGAGAGCGCGGTGGAACAGAGAGCT 253
DB 1183 SerGlyArgLeuSerLeuProAlaLeuSerProAspSerGlySer----- 1197
QY 254 CCGCAGACAGCCGAGCGCGCGCGAGAGAGCGCGCAGCGCATGAGAGGTCCGAG 313
DB 1198 ArgAspSerArgSerProAspAlaAspAlaAsnArgMetIleAspIleGluGlyUaAsp 1217
QY 314 CCGAGAACATG-----CAGCCAGCAACTCGG-----CGACAGAGTTCAAGCTG 358
DB 1218 SerGluSerGlnAspSerAspGlnProGlyPheArgAsnArgThrPheSerPro 1237
QY 359 TTGACAGTGAGAGAGCTGGAAGTGTTCGACACACTCAATACCTGATGCGCACA 418
DB 1238 GluGlnLeuAspGluLeuGluGlyGluPheAspGlySerHisTyProCysValAsnThr 1257
QY 419 AGAAGGAACTTCCGAAACTTAAGTGTGACTGAAGACAAAGTGGGTTGTTAAG 478
DB 1258 ArgGluTyLeuAlaAlaArgThrAlaLeuSerGluAlaArgValGlnValTTPPheSer 1277

QY 479 AATAAAAGCCAGATGTGCGCATCAGAGA 511
DB 1278 AsnArgArgAlaTyTTPArgArgHisGlnArg 1288

RESULT 8
US-10-221-625-88
; Sequence 88, Application US/10221625
; Publication No. US20040033942A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: YUE, Henry
; APPLICANT: LAL, Preeti
; APPLICANT: LU, Dying Aina M.
; APPLICANT: PATTERSON, Chandra
; APPLICANT: AZIMZAI, Valda
; APPLICANT: BANDMAN, Olga
; APPLICANT: TANG, Y. Tom
; APPLICANT: MATHUR, Preete
; APPLICANT: SHAH, Purvi
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: REDDY, Roopa
; TITLE OF INVENTION: TRANSCRIPTION FACTORS
; FILE REFERENCE: PF-0761 PCT
; CURRENT APPLICATION NUMBER: US/10/221,625
; CURRENT FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 214
; SOFTWARE: PERL Program
; SEQ ID NO 88
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040033942A1 3717139CD1
US-10-221-625-88

Alignment Scores:
Pred. No.: 1.07e-06 Length: 215
Score: 187.50 Matches: 37
Percent Similarity: 58.89% Conservative: 16
Best Local Similarity: 41.11% Mismatches: 30
Query Match: 17.86% Indels: 7
DB: 12 Gaps: 1

```

US-09-867-753-1 (1-577) x US-10-221-625-88 (1-215) -----CAGGCGCCATGAGAGT 307
QY 269 CCCCCCGAGAGAGCCGCGCC-----
Db 9 ProProProValSerProGluLeuLysAspArgLysGluAspAlaLysGlnAsp 28
QY 308 CCGCAGCCCGAGAGAGCCAGCCAGCAACTCGGCGCAGAGAAAGTTCAAGCTGTTGACAGGTG 367
Db 29 GluGluGlnThrLysLeuLysGlnArgSerArgThrAspMetThrLeuGlnGlnLeu 48
QY 368 GAGAGACTGGAAGAGTGTTCACACACTCACTCAATACCTGATGTGCCCAAGAGAGAA 427
Db 49 AsnGluLeuGluArgLeuPheAspGluThrHisTyrProAspAlaPheMetArgGlnGln 68
QY 428 CTTCGCGAAACTTGAAGTGTGCTGAGTGAAGCAAGATCGGGTGTGTTAAGATAAAG 487
Db 69 LeuSerGlnArgLeuLysLeuSerGlnAlaArgValGlnValTyrPheGlnAsnArgArg 88
QY 488 GCCAGATGTAGGCGCAGCATCAGAGAGAAATTA 517
Db 89 AlaLysCysArgLysGlnGlnLeu 98
RESULT 9
US-10-301-822-165
; Sequence 165, Application US/10301822
; Publication No. US20030148410A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: BURGART, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MEMO1-029P2RNM
; CURRENT APPLICATION NUMBER: US/10/301,822
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 165
; LENGTH: 217
; TYPE: PRN
; ORGANISM: Homo Sapiens
US-10-301-822-165
Alignment Scores:
Pred. No.: 1,86e-06 Length: 217
Score: 184.50 Matches: 49
Percent Similarity: 44.97% Conservative: 27
Best Local Similarity: 28.99% Mismatches: 54
Query Match: 17.57% Indels: 39
DB: 14 Gaps: 5
US-09-867-753-1 (1-577) x US-10-301-822-165 (1-217)
QY 113 TCAGAGCGCAGAGAGCCATGTTGGCCAGAGAGCTCCAGGCTCATGGGTAAATGAACCTT 172
Db 2 ThrSerSerTyrGlnHisValLeuGluArgGlnProAlaLeuGlnLysArgLeuAspSer 21
QY 173 GAGGCG-----GGTGTGAACCAAGAGAGCGGC 159
Db 22 ProGlyAsnLeuAspThrLeuGlnAlaLysLysAsnPheSerValSerHisLeuLeuAsp 41

QY 200 ATGAACCGCATGCGCGCATGATCC-----GAGGCGGC 235
Db 42 LeuGluGluAlaGlyAspMetValAlaAlaGlnAlaAspGluAsnValGlnGlnAlaGly 61
QY 236 -----GGTGAACCAAGAGAGCGCGCAGCGCG 265
Db 62 ArgSerLeuLeuGlnSerProGlyLeuThrSerGlySerAspThrProGlnGlnAsp--- 80
QY 266 CAGCCCCCGGAGAGAGCGCCGCGCCATGAGAGAGTCCGACGCCCGAGAACATG 325
Db 81 -----AsnAspGlnLeuAsnSerGlnGluLysLysAspArg 92
QY 326 CAGCCACGAATCTCGCGCAGAGATTCAAGCTTTCAGAGTGAAGAGAGCTGGAAGTGT 385
Db 93 LysGlnArgArgAsnArgThrThrPheAsnSerSerGlnLeuGlnAlaLeuGlnArgVal 112
QY 386 TTCGACACACTCAATACCTGATGCGCCACAGAGAGAACTTGCCGAAACTTAGGT 445
Db 113 PheGluArgThrHisTyrProAspAlaPheValArgGluAspLeuAlaArgArgValAsn 132
QY 446 GTGACTGAAGACAAAGTGGCTTGTGTTAAGATAAAGGCGCAGATGTAGCGCAGCAT 505
Db 133 LeuThrGluAlaArgValGlnValTyrPheGlnAsnArgArgAlaLysSerHeArgArgAsn 152
QY 506 CAGAGAGAAATTAATGCTCGCCCAATGAA 532
Db 153 GluArg---AlaMetLeuAlaAsnLys 160
RESULT 10
US-10-301-822-167
; Sequence 167, Application US/10301822
; Publication No. US20030148410A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: BURGART, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MEMO1-029P2RNM
; CURRENT APPLICATION NUMBER: US/10/301,822
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 167
; LENGTH: 245
; TYPE: PRN
; ORGANISM: Homo Sapiens
US-10-301-822-167
Alignment Scores:
Pred. No.: 1,88e-06 Length: 245
Score: 184.50 Matches: 49
Percent Similarity: 44.97% Conservative: 27
Best Local Similarity: 28.99% Mismatches: 54
Query Match: 17.57% Indels: 39
DB: 14 Gaps: 5
US-09-867-753-1 (1-577) x US-10-301-822-167 (1-245)
QY 113 TCAGAGCGCAGAGAGCCATGTTGGCCAGAGAGCTCCAGGCTCATGGGTAAATGAACCTT 172
Db 2 ThrSerSerTyrGlnHisValLeuGluArgGlnProAlaLeuGlnLysArgLeuAspSer 21

```

Db      2  ThrsSerIeTYGLYHIEValleuGIuAArgInPrOAlaleuGIyGLyAArgLeuAePser 21
QY      173 GAGGGC-----GGTGAACCCAGAGAACGCC 199
           |||
Db      22 ProGIaAnleuAspThrleuGlnAlaIlybLyAsnPheserValSerHisleuLeuAsp 41
QY      200 ATGAACCGCATGCGCGCATGATCC-----GAGCGCGGC 239
           ::  |||  ::::
Db      42 LeuGIuGIuAlaGIYAspMetValAlaIAlaGlnAlaAspGIuAnValGIyGIuAlaGIy 61
QY      236 -----GGTGAACCCAGAGACCTCGGCAGCAGCCG 265
           |||
Db      62 ArgSerIeLeuGIuSerProGIyLeuThrSerGIySerAspThrProGIuGlnAsp--- 80
QY      266 CAGCCCCCGCGAGAGAGCGCGCCAGCGCGCATGAGAGGTATCCGACCCCGAGAACATG 325
           ::::
Db      81 -----AsnAspGIuLeuAsnSerGIuGIuIyIyIyIyIyArg 92
QY      326 CAGCCACGAACTCGCGCGCAGAAAGTTCCAGCTGTTCAGGTGAGAGAGCTGAAAGTGT 385
           ::  |||  |||  |||  ::::
Db      93 LysGIuAlaArgAsnArgThrThrPheAsnSerSerGIuLeuGlnAlaIeGIuAlaArgVal 112
QY      386 TTCGACACACTCAATPACCTGATGTGGCCOACAGAAAGGAACTTGCGCGAAACCTTAGT 445
           |||  |||  |||  |||  |||  ::::
Db      113 PheGIuAlaGIuThHisIeTYrProAspAlaPheValArgGIuAspLeuAlaArgValAsn 122
QY      446 GTGACTGAAGACAAAGTGCGGGGTGGTTGTTAAGATAAAGGCGCAGATGTAGCGCAT 505
           ::  |||  ::  |||  ::::  ::::  ::::  |||  |||
Db      133 LeuThrGIuAlaIaArgValGIuAlaITrpPheGlnAsnArgArgAlaIyIyPheArgArgAsn 152
QY      506 CAGAGAGAAATTATGCTCGCCCAATGA 532
           ::::  |||  |||  |||  |||  |||  ::::
Db      -153 GIuArg---AlaMetIeLeuAlaAsnIy 160

```

```

RESULT 11
US-10-158-160A-16
Sequence 16, Application US/10158160A
Publication No. US20030059805A1
GENERAL INFORMATION:
APPLICANT: RAPPOLD-HOERNBRAND, GUDDRUN
APPLICANT: RAO, ERCOLE
FILE OF INVENTION: HUMAN GROWTH GENE AND SHORT STATURE GENE REGION
FILE REFERENCE: 108351-00004
CURRENT APPLICATION NUMBER: US/10/158,160A
CURRENT FILING DATE: 2002-08-20
PRIOR APPLICATION NUMBER: 09/147,699
PRIOR FILING DATE: 1999-06-24
PRIOR APPLICATION NUMBER: PCT/EP97/05355
PRIOR FILING DATE: 1997-03-29
PRIOR APPLICATION NUMBER: 60/027,633
PRIOR FILING DATE: 1996-10-01
PRIOR APPLICATION NUMBER: EP/97100583.0
PRIOR FILING DATE: 1997-01-16
NUMBER OF SEQ ID NOS: 55
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 16
LENGTH: 190
TYPE: PRT
ORGANISM: Homo sapiens
US-10-158-160A-16

Alignment Scores:
Pred. No.: 6,09e-06 Length: 190
Score: 178.00 Matches: 33
Percent Similarity: 67.12% Conservativity: 16
Best Local Similarity: 45.21% Mismatches: 24
Query Match: 16.95% Indels: 0
DB: 14 Gaps: 0

```

US-09-867-753-1 (1-577) x US-10-158-160A-16 (1-190)

QY 299 ATGGAGGTCGCGACGCCGAGAACATGTCGAGCGACGAAATTCACGCTG 358
||||| :||||| |||||||

[illegible]

```

RESULT 12
US-10-158-160A-13
: Sequence 13, Application US/10158160A
: Publication No. US20030059805A1
: GENERAL INFORMATION:
: APPLICANT: RAPPOLD-HOERBRAND, GUDRUN
: APPLICANT: RAO, ERICOLE
: TITLE OF INVENTION: HUMAN GROWTH GENE AND SHORT STATURE GENE REGION
: FILE REFERENCE: 108351-00004
: CURRENT APPLICATION NUMBER: US/10/158,160A
: CURRENT FILING DATE: 2002-08-20
: PRIOR APPLICATION NUMBER: 09/147,699
: PRIOR FILING DATE: 1999-06-24
: PRIOR APPLICATION NUMBER: PCT/EP97/05355
: PRIOR FILING DATE: 1997-09-29
: PRIOR APPLICATION NUMBER: 60/027,633
: PRIOR FILING DATE: 1996-10-01
: PRIOR APPLICATION NUMBER: EP/97100583.0
: PRIOR FILING DATE: 1997-01-16
: NUMBER OF SEQ ID NOS: 55
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 13
: LENGTH: 225
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-10-158-160A-13

```

Alignment Scores:			
Pred. No.:	8,14e-06	Length:	225
Score:	176.50	Matches:	49
Percent Similarity:	49.71%	Conservative:	37
Best Local Similarity:	28.32%	Mismatches:	70
Query Match:	16.81%	Indels:	17
	14	Gaps:	4

QY	US-09-867-753-1 (1-577) x US-10-158-160A-13 (1-225)	
44	GACACGGTGTTCACGCTGAGTGTATCCACAGGTAAAAATAAGCCCAACACTGAGTGG	10
Db	31 AbspSer1ethrlyr---ArggluValLeuGluSerGlyLeuAlaMetArgSerGluLeu	49
QY	104 GGGGACATCAAGCCGACAGAGCCCATGTGGCCAAAGACCCGAGGCTTCATGGTAT	16
Db	50 GlyThSerApsSerSerLeuGlnApsr1IethrGluGlyGlyNIeCyproValHis	69
QY	164 ATGAACCTCGAGGGCGGCTGTGAACCAACGAGAACGCAATGACCCGGATGCG---	22
Db	70 LeuPheLysApsrIleValAspMetAspGlyGlyIleuValGluPheIleThrAlaArg	89
QY	221 ATCCCGAGAGCGCGCGGTGAAAAACGAGAGCTTGCGACAGCCGACCCCGCGGAG	28
Db	90 ValAlaGluGlyIleTyrGluCyArgSerGluIlyAsrGluApsrValIysSerGluApsrGlu	10
QY	281 GAGCCGAGCCAGCGCGCATGAGAGGCTCCGACGCCCGAAGAACATGACACGAACTGG	34
Db	110 Asp-----GlyGlnThrIlyIleuValGluMetArgSer	12
QY	341 CGCAGAAATTCACGCTGTTCAGAGTGCAGAGAGCTGAAAAAGTGTTCGACACATCA	40

Db 121 ArgThrAsnPhenThreLeuGlnLeuAangluleGluArgLeuPheAspGluThrHis 140
 QY 401 TACCCTGATGTCGCCACAAGAGGAAGCTTGCCGAAACTTAGAGTGTGACTGAGACAAA 460
 Db 141 TyrProAspAlaPheMetArgGlnGluLeuSerGlnArgLeuGlnYleuSerGlnAlaArg 160
 QY 461 GTCCGGGTTGGTTTAAGAAATAAAGGCGCAATGTAGCGCATATAGAGAGA----- 514
 Db 161 ValGlnValIleTyrPheGlnAsnArgAlaIleCysArgGlnGlnGlnGlnMetHis 180
 QY 515 -----TTAATGCTCGCCCATGAACTACGTGCT 541
 Db 181 LysGlyValIleLeuGlnYThrAlaAsnHisLeuAspAla 193

RESULT 13

US-10-158-160A-11
 ; Sequence 11, Application US/10158160A
 ; Publication No. US20030059805A1
 ; GENERAL INFORMATION:
 ; APPLICANT: RAPFOLD-HOERBRAND, GUDRUN
 ; APPLICANT: RAO, ERCOLE
 ; TITLE OF INVENTION: HUMAN GROWTH GENE AND SHORT STATURE GENE REGION
 ; FILE REFERENCE: 108351-00004
 ; CURRENT APPLICATION NUMBER: US/10158, 160A
 ; CURRENT FILING DATE: 2002-08-20
 ; PRIOR APPLICATION NUMBER: 09/147, 699
 ; PRIOR FILING DATE: 1999-06-24
 ; PRIOR APPLICATION NUMBER: PCT/EP97/05355
 ; PRIOR FILING DATE: 1997-09-29
 ; PRIOR APPLICATION NUMBER: 60/027, 633
 ; PRIOR FILING DATE: 1996-10-01
 ; PRIOR APPLICATION NUMBER: EP/97100583.0
 ; PRIOR FILING DATE: 1997-01-16
 ; NUMBER OF SEQ ID NOS: 55
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 11
 ; LENGTH: 292
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-10-158-160A-11

Alignment Scores:

Pred. No.: 8,31e-06 Length: 292
 Score: 176.50 Matches: 49
 Percent Similarity: 49.71% Conservative: 37
 Best Local Similarity: 28.32% Mismatches: 70
 Query Match: 16.81% Indels: 17
 DB: 14 Gaps: 4

US-09-867-753-1 (1-577) x US-10-158-160A-11 (1-292)

QY 44 GACACGCTGTTTACGCGCTGATATACAGGTAAATATAGCCCACTCAGCTG 103
 Db 31 AspSerIleThrTyr---ArgGluValLeuGlnSerGlnYleuAlaArgSerArgGluLeu 49
 QY 104 GGGGCGACATCAAGCGCAGAGGCGCATGTTGGCCAGAGAGCTCAGGCTCATGGGTAAAT 163
 Db 50 GlnThrSerAspSerIleuGlnAspIleThrGlnGlnGlnYleuGlnCysProValHis 69
 QY 164 ATGAACCTTAGAGGCGGCTGTGAACACAGAGACGCGATGAACCGCATGCG---GGCANG 220
 Db 70 LeuPheIleAspHisValAspAsnAspLysGlnLysLeuLysGlnPheGlnThrAlaArg 89
 QY 221 ATCCCGAGGCGCGCGGTGGAACACAGAGCTCGGAGACGCGGAGCCCGCGGAG 280
 Db 90 ValAlaGlnIleYleTyrGlnCysLysGlnLysArgGlnValLysSerGlnAspGln 109
 QY 281 GAGCCGCGCGCGCGCGCATGAGGGTCCGACCGCGAGACATGAGCCAGCACTCGG 340
 Db 110 Asp-----GlnGlnThrLysLeuLysGlnArgArgSer 120
 QY 341 CGCAGAGAGTTACGCTGTTGACGTTGAGAGAGCTGGAAGGTGTTTCCGACACTCAA 400

Db 121 ArgThrAsnPhenThreLeuGlnLeuAangluleGluArgLeuPheAspGluThrHis 140
 QY 401 TACCCTGATGTCGCCACAAGAGGAAGCTTGCCGAAACTTAGAGTGTGACTGAGACAAA 460
 Db 141 TyrProAspAlaPheMetArgGlnGluLeuSerGlnArgLeuGlnYleuSerGlnAlaArg 160
 QY 461 GTCCGGGTTGGTTTAAGAAATAAAGGCGCAATGTAGCGCATATAGAGAGA----- 514
 Db 161 ValGlnValIleTyrPheGlnAsnArgAlaIleCysArgGlnGlnGlnGlnMetHis 180
 QY 515 -----TTAATGCTCGCCCATGAACTACGTGCT 541
 Db 181 LysGlyValIleLeuGlnYThrAlaAsnHisLeuAspAla 193

RESULT 14

US-10-087-192-1983
 ; Sequence 1983, Application US/10087192
 ; Publication No. US20020182586A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Morris, David W.
 ; APPLICANT: Engelhard, Eric K.
 ; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
 ; TITLE OF INVENTION: CANCER
 ; FILE REFERENCE: 529452000122
 ; CURRENT APPLICATION NUMBER: US/10/087, 192
 ; CURRENT FILING DATE: 2002-03-01
 ; PRIOR APPLICATION NUMBER: US 09/747, 377
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: US 09/798, 586
 ; PRIOR FILING DATE: 2001-03-02
 ; NUMBER OF SEQ ID NOS: 2059
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1983
 ; LENGTH: 489
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 ; FEATURE:
 ; NAME/KEY: VARIANT
 ; LOCATION: (1)...(489)
 ; OTHER INFORMATION: Xaa = Any Amino Acid
 ; US-10-087-192-1983

Alignment Scores:

Pred. No.: 8,67e-06 Length: 489
 Score: 176.50 Matches: 47
 Percent Similarity: 49.22% Conservative: 16
 Best Local Similarity: 36.72% Mismatches: 44
 Query Match: 16.81% Indels: 21
 DB: 12 Gaps: 4

US-09-867-753-1 (1-577) x US-10-087-192-1983 (1-489)

QY 173 GGGGCGGTGTGACACAGAGAGGACGATGACCGC-----GATGCGGG 217
 Db 169 GlnGlyAspLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 188
 QY 218 ATGATCCCGAGGCGCGCGGTGGAACACAGAGCTCGGACGAGCGCGCGCGCG 277
 Db 189 Leu-----GlnAspLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 196
 QY 278 GAGAGCCCGCGCGCGCGCATGAGGGTCCGACCGCGAGACATGACACCGAGACT 337
 Db 197 AspGlnGlySerAspValGlnSerGlnProAspLeuProLeuLysArgLysGlnArg 216
 QY 338 CGGCGAGCAAGATTCACGCTGTTGACGAGGAGAGAGAGAGAGAGAGAGAGAGAG 397
 Db 217 SerArgThrThrPheThrAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 236
 QY 398 CAATACCTGATGTGCCACAAGAGGAAGCTTGCCGAAACTTAGAGTGTGACTGAGAC 457
 Db 237 HisTyrProAspLysIleThrArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 256
 QY 458 AAGTGGGGTGGTTTAAGAAATAAAGGCGCAATGTAGCGCATCAGAGAGAAATTA 517

Db 257 ArgValGlnValTPrpSerAsnArgAlaArgTPrpArgLysGlnAlaGly----- 274

QY 518 ATGCTCGCCATGAACCTACGTGCT 541

Db 275 -----AlaAsnGlnLeuAlaAla 280

RESULT 15

US-10-245-171A-7

; Sequence 7, Application US/10245171A

; Publication No. US20030124102A1

; GENERAL INFORMATION:

; APPLICANT: Rudnicki, Michael A.

; APPLICANT: Seale, Patrick

; TITLE OR INVENTION: Pax-Encoding Vector and Use Thereof

; FILE REFERENCE: 50120/006002

; CURRENT APPLICATION NUMBER: US/10/245,171A

; CURRENT FILING DATE: 2003-03-11

; PRIOR APPLICATION NUMBER: US 60/322,923

; PRIOR FILING DATE: 2001-09-17

; NUMBER OF SEQ ID NOS: 13

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO: 7

; LENGTH: 479

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-245-171A-7

Alignment Scores:

Pred. No.:

Score: 9,496-06

Percent Similarity: 176.00

Best Local Similarity: 56.04%

Query Match: 42.86%

DB: 16.76%

Gaps: 14

Length: 479

Matches: 39

Conservative: 12

Mismatches: 38

Indels: 2

Gaps: 1

US-09-867-753-1 (1-577) x US-10-245-171A-7 (1-479)

QY 269 CCCCCCGGAGAGACCCGCGCCAGCGGCGCATGAGCGTCCGACCCCGAAGACATGCAG 328

Db 199 ProGlnSerAspGlnGlySerAspIleAspSerGlnProAspLeuProLeuLysArgLys 218

QY 329 CCAGCACTCGGCGGACGAGGAGTTCAGCTGTGACAGAGCTGAGAGCTGGAAGTGTTC 388

Db 219 GlnArgArgSerArgThrPheThrAlaGlnGlnLeuGlnLeuGlnAlaPhe 238

QY 389 CGACACACTCATATACCTGATGTGCCACAGAGGAACTGCCGAAAATTAGGTGTG 448

Db 239 GlnArgThrHisTyrProAspIleTyrThrArgGlnGlnLeuAlaGlnAlaLysLeu 258

QY 449 ACTGAAGACAAAGTGGGGTTGGTTTAAGATTAAGAGGCGCAGATGAGCGACAT--- 505

Db 259 ThrGlnAlaArgValGlnValTPrpSerAsnArgAlaArgTPrpArgLysGlnAla 278

QY 506 ---CAGAGAGATTATGCTGCGCAATGAACCTA 535

Db 279 GlyAlaAsnGlnLeuMetAlaPheAsnHisLeu 289

Search completed: June 16, 2004, 19:05:52
Job time : 49.5 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 16, 2004, 18:58:15 ; Search time 19.5 Seconds
(without alignments)
3055.198 Million cell updates/sec

Title: US-09-867-753-1
Perfect score: 1050
Sequence: 1 tccacatcagcgctccag.....tctacatcgctgtagactag 577

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues
Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame_n2p.model -DEV=xlp
-O=/cgn2_1/USPTO_spool_p/US09867753/runat_16062004_163636_27826/app_query.fasta_1.775
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=xrai -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09867753_@CGN1_1.1.27_@runat_16062004_163636_27826 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA:*
1: /cgn2_6/pdata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/pdata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/pdata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/pdata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/pdata/2/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/pdata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	190.5	18.1	302	4	US-08-957-351-7
2	190.5	18.1	313	4	US-08-957-351-9
3	189.5	18.0	302	4	US-08-957-351-30
4	188	17.9	302	4	US-08-957-351-3
5	183.5	17.5	305	2	US-08-775-009-37
6	181.5	17.3	315	4	US-08-957-351-27
7	176	16.8	247	3	US-09-129-888-2
8	174	16.5	234	4	US-09-105-470B-4
9	173	16.5	185	4	US-09-976-594-443
10	171.5	16.3	289	4	US-09-976-594-945
11	169.5	16.1	271	4	US-08-957-351-26
12	169	16.1	241	4	US-09-636-215-842

13	169	16.1	241	4	US-09-685-166A-842	Sequence 842, App
14	167	15.9	60	4	US-08-754-477A-18	Sequence 18, Appl
15	167	15.9	60	4	US-08-754-477A-19	Sequence 19, Appl
16	164.5	15.7	132	4	US-08-754-477A-135	Sequence 135, App
17	164.5	15.7	271	4	US-08-754-477A-2	Sequence 2, Appl
18	164.5	15.7	271	4	US-08-754-477A-5	Sequence 5, Appl
19	163	15.5	60	4	US-08-754-477A-11	Sequence 11, Appl
20	162	15.4	60	4	US-08-754-477A-10	Sequence 10, Appl
21	162	15.4	234	4	US-09-105-470B-2	Sequence 2, Appl
22	161	15.3	60	4	US-08-754-477A-14	Sequence 14, Appl
23	161	15.3	240	4	US-09-636-735A-2	Sequence 2, Appl
24	161	15.3	240	4	US-09-636-735A-12	Sequence 12, Appl
25	160	15.2	60	2	US-08-775-009-38	Sequence 38, Appl
26	160	15.2	60	4	US-08-754-477A-13	Sequence 13, Appl
27	160	15.2	60	4	US-08-754-477A-15	Sequence 15, Appl
28	160	15.2	60	4	US-08-754-477A-17	Sequence 17, Appl
29	160	15.2	60	4	US-08-754-477A-20	Sequence 20, Appl
30	160	15.2	99	4	US-09-031-962D-24	Sequence 24, Appl
31	159	15.1	60	4	US-08-754-477A-16	Sequence 16, Appl
32	159	15.1	436	2	US-08-958-642-4	Sequence 4, Appl
33	159	15.1	436	3	US-08-778-394-2	Sequence 2, Appl
34	159	15.1	436	3	US-08-778-423A-4	Sequence 4, Appl
35	158.5	15.1	271	4	US-08-754-477A-121	Sequence 121, App
36	158.5	15.1	271	4	US-08-754-477A-125	Sequence 125, App
37	158	15.0	60	4	US-08-754-477A-7	Sequence 7, Appl
38	157.5	15.0	271	4	US-08-754-477A-131	Sequence 131, Appl
39	157	15.0	60	4	US-08-754-477A-12	Sequence 12, Appl
40	157	15.0	60	4	US-08-754-477A-23	Sequence 23, Appl
41	156.5	14.9	302	2	US-08-203-532F-4	Sequence 4, Appl
42	156.5	14.9	302	3	US-08-950-860-16	Sequence 16, Appl
43	156.5	14.9	302	3	US-09-078-465-4	Sequence 4, Appl
44	156.5	14.9	302	5	PCT-US95-01882A-4	Sequence 4, Appl
45	155	14.8	434	2	US-08-710-249-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-957-351-7
Sequence 7, Application US/08957351
Patent No. 6306586
GENERAL INFORMATION:
APPLICANT: Murray, Elena
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESSES:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109-2170
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/957,351
FILING DATE: 24-OCT-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: UIA-024.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 302 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-957-351-7

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
4.23e-10	302	49	23	45	25	5
Score: 190.50						
Percent Similarity: 50.70%						
Best Local Similarity: 34.51%						
Query Match: 18.14%						

US-09-867-753-1 (1-577) x US-08-957-351-7 (1-302)

QY 98 CAGCTGGGGGAGCATCAAGCGGAGGCGCATGTTGGCCAAAGAGCTCCAGGCTCATG 157
Db 2 GIUPHegilyLeuLeuSerGIuAgluA1aArg-----SerProAlaLeuSer 17
QY 158 GGTAAATATGAACCCCTGAGGGCGGTGTGAACACAGAGAAGCGCATGAACCGCATGCGCGC 217
Db 18 LeuSer-----AspAlaGlyThrProHisPro----- 26
QY 218 ATGATCCCGAGGCGCGCGGTGGAACACAGAGCTCCGAGAGCGCGAGCCCGCGCG 277
Db 27 GlnLeuProGluHisGlyCysIysGlyGlnGluHisSerAsp----- 40
QY 278 GAGAGCGCGCGCGAGGCGCATGAGAGGTCGCGAGCGCGAGAAC-----ATG 325
Db 41 SerGluIyAlaSerIaSerLeuProGlyGlySerProGluAspIySerLeuIySlyS 60
QY 326 CAGCCAGCACTCGGCGGCGGAGGTTCAAGCTGTTGCAAGTGAAGAGTGAAGTGT 385
Db 61 LysGlnArgArgGlnArgThrHisPheThrSerGlnGlnLeuGlnGluLeuA1aTrpHis 80
QY 386 TTCGACACACTCAATACCTGATGTGTCGCCACAAGAAGGAACTTGCAGAACTTAGGT 445
Db 81 PheGlnArgAsnArgIyTrpProAspMetSerThrArgGlnGluLeuA1aValTrpHis 100
QY 446 GTGACTGAAGACAAAGTGGTTGTTTAAGATTAAGGAGGCGAGATGAGCGACAT 505
Db 101 LeuThrGluAlaArgValaArgValaTrpPheIyAsnArgArgAlaIyStrpArgIySarg 120
QY 506 CAGAGA 511
Db 121 GluArg 122

RESULT 2
US-08-957-351-9
Sequence 9, Application US/08957351
Patent No. 6306586
GENERAL INFORMATION:
APPLICANT: Semina, Elena
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF CATARACTS
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESSES:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109-2170
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/957,351
FILING DATE: 24-OCT-1997
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: UIA-024.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000

INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 313 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-957-351-9

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
4.27e-10	313	49	23	45	25	5
Score: 190.50						
Percent Similarity: 50.70%						
Best Local Similarity: 34.51%						
Query Match: 18.14%						

US-09-867-753-1 (1-577) x US-08-957-351-9 (1-313)

QY 98 CAGCTGGGGGAGCATCAAGCGGAGGCGCATGTTGGCCAAAGAGCTCCAGGCTCATG 157
Db 2 GIUPHegilyLeuLeuSerGIuAgluA1aArg-----SerProAlaLeuSer 17
QY 158 GGTAAATATGAACCCCTGAGGGCGGTGTGAACACAGAGAAGCGCATGAACCGCATGCGCGC 217
Db 18 LeuSer-----AspAlaGlyThrProHisPro----- 26
QY 218 ATGATCCCGAGGCGCGCGGTGGAACACAGAGCTCCGAGAGCGCGAGCCCGCGCG 277
Db 27 GlnLeuProGluHisGlyCysIysGlyGlnGluHisSerAsp----- 40
QY 278 GAGAGCGCGCGCGAGGCGCATGAGAGGTCGCGAGCGCGAGAAC-----ATG 325
Db 61 LysGlnArgArgGlnArgThrHisPheThrSerGlnGlnLeuGlnGluLeuA1aTrpHis 80
QY 386 TTCGACACACTCAATACCTGATGTGTCGCCACAAGAAGGAACTTGCAGAACTTAGGT 445
Db 81 PheGlnArgAsnArgIyTrpProAspMetSerThrArgGlnGluLeuA1aValTrpHis 100
QY 446 GTGACTGAAGACAAAGTGGTTGTTTAAGATTAAGGAGGCGAGATGAGCGACAT 505
Db 101 LeuThrGluAlaArgValaArgValaTrpPheIyAsnArgArgAlaIyStrpArgIySarg 120
QY 506 CAGAGA 511
Db 121 GluArg 122

RESULT 3

US-08-957-351-30
Sequence 30, Application US/08957351
Patent No. 6306586
GENERAL INFORMATION:
APPLICANT: Semina, Elena
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF CATARACTS
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESSES:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA


```

; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,351
; FILING DATE: 24-OCT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: UIA-024.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 302 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-08-957-351-30
;
; Alignment Scores:
; Pred. No.: 5,27e-10 Length: 302
; Score: 189.50 Matches: 49
; Percent Similarity: 50.00% Conservative: 22
; Best Local Similarity: 34.51% Mismatches: 46
; Query Match: 18.05% Indels: 25
; DB: Gaps: 5
;
; US-09-867-753-1 (1-577) x US-08-957-351-30 (1-302)
;
; QY 98 CAGCTGGGGGAGCATCAAGCCAGAGGCGCATGTTGCCAAGAGCTCCAGGCGCTCATG 157
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; DB 2 GluphegilyleuleuSerGluAlaValArg-----ThrpAlaLeuSer 17
;
; QY 158 GGTAAATATGAAACCTGAGGCGGTGTGAACCAAGAGAGCGCATGAACCGCATGGCGGC 217
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; DB 18 LeuSer-----AspAlaGlyThrProHisPro----- 26
;
; QY 218 ATGATCCCGGAGGGGCGGTGGAACAGAGAGCGCTCGGAGAGCGCGAGCCCGCGCG 277
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; DB 27 GluLeuProGluHisGlyCysValGlyGlnGluHisSerAsp----- 40
;
; QY 278 GAGGAGCGGCGCCAGGCGCGCATGAGAGGTCGCGAGCCCGAGAAC-----ATG 325
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; DB 41 SerGluValAlaSerAlaSerLeuProGlyGlySerProGluAspGlySerLeuValys 60
;
; QY 336 CAGCCACGAACTCGGCGCAGAAATTACGCTGTTCAGAGTGAAGAGCTGGAAGTT 385
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; DB 61 LysGlnArgArgGlnArgThrHisPheThrSerGlnGlnLeuGlnLeuGlnAlaThr 80
;
; QY 386 TTCGACACACTCAATACCTGATGTCGCCACAGAGGAAGCTGCCGAAAATTAGGT 445
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; DB 81 PheGlnArgAsnArgGlyTrpAspMetSerThrArgGlnGlnIleAlaValTrpThrAsn 100
;
; QY 446 GTGACTGAGAGCAAGTGGGTTGTTTAAAGTAAGAGGCGCATGTAGCGGACAT 505
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; DB 101 LeuThrGluAlaArgValArgValTrpPheLysAsnArgArgAlaLysTrpArgLysArg 120
;
; QY 506 CAGAGA 511
;   ||| |||
; DB 121 GlnArg 122
;
; RESULT 4
; US-08-957-351-3
; Sequence 3, Application US/08957351
; Patent No. 6306586
; GENERAL INFORMATION:
; APPLICANT: Semina, Elena

```

```

; APPLICANT: Murray, Jeffrey C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF CATARACTS
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HONG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,351
; FILING DATE: 24-OCT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: UIA-024.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 302 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-08-957-351-3
;
; Alignment Scores:
; Pred. No.: 7.32e-10 Length: 302
; Score: 188.00 Matches: 41
; Percent Similarity: 59.41% Conservative: 19
; Best Local Similarity: 40.59% Mismatches: 31
; Query Match: 17.90% Indels: 10
; DB: Gaps: 2
;
; US-09-867-753-1 (1-577) x US-08-957-351-3 (1-302)
;
; QY 221 ATCCCGAGGGGCGGCGGTGGAACAGAGGCTCGGAGAGCGCGAGCCCGCGCGAG 280
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; DB 28 LeuProGluHisGlyCysValGlyGlnGluHisSerAsp-----Ser 41
;
; QY 281 GAGCGGCGCCAGGCGCGCATGAGAGGTCGCGAGCCCGAGAAC-----ATGAG 328
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; DB 42 GlnValAlaSerAlaSerLeuProGlyGlySerProGluAspGlySerLeuValys 61
;
; QY 329 CCAGCAACTCGGCGCAGCAAGTTCAAGCTGTTCAGAGTGAAGAGCTGGAAGTTTTC 388
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; DB 62 GlnArgArgGlnArgThrHisPheThrSerGlnGlnLeuGlnLeuGlnAlaThrPhe 81
;
; QY 389 CGACACACTCAATACCTGATGTCGCCACAGAGGAAGCTGCCGAAAATTAGGTG 448
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; DB 82 GlnArgAsnArgGlyTrpAspMetSerThrArgGlnGlnIleAlaValTrpThrAsn 101
;
; QY 449 ACTGAAGCAAGTGGGTTGTTTAAAGTAAGAGGCGCATGTAGCGGACATGAG 508
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; DB 102 ThrGluAlaArgValArgValTrpPheLysAsnArgArgAlaLysTrpArgLysArg 121
;
; QY 509 AGA 511
;   ||| |||
; DB 122 Arg 122
;
; RESULT 5
; US-08-975-009-37
; Sequence 37, Application US/08775009
; Patent No. 5935783

```



```

QY 328 GCCACGAACTCGCGCAGCAAGTTCACGCTGTTCAGGTGAGAGACCTGGAAGTGTCTT 387
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 89 66lnrgrgaglnrgrglnrgrhlspherrhserrgrnglnrlnrlnrlnrlnrlnrlnr 109
QY 388 CCGACACACTCAATACCTCGATGTGCTCCCAAGAGAGAACTTGGCAAACTTGAAGTGT 447
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109 66lnrgrgaglnrgrglnrgrhlspherrhserrgrnglnrlnrlnrlnrlnrlnr 129
QY 448 GACTGAGCAAGCAAGCGGCTTGTTCAGTATTAAGATTAAGCAAGTATTAAGCAAGTAT 507
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 129 unhrgrluprgrgrglnrgrglnrgrhlspherrhserrgrnglnrlnrlnrlnrlnr 149
QY 508 GAGAGATTAATGCTC 523
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 149 unhrgrglnrlnrlnr 154

```

RESULT 7

```

US-09-129-888-2
; Sequence 2, Application US/09129888B
; Patent No. 6063912
; GENERAL INFORMATION:
; APPLICANT: CHUN, Jong Yoon
; APPLICANT: HAN, Yun Jeong
; TITLE OF INVENTION: Placenta trophoblast-specific gene
; FILE REFERENCE: 1942/29
; CURRENT APPLICATION NUMBER: US/09/129,888B
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: WordPerfect 6.1 Windows
; SEQ ID NO 2
; LENGTH: 247
; TYPE: PRT
; ORGANISM: mouse
US-09-129-888-2

```

```

Alignment Scores:
Pred. No.: 9,71e-09 Length: 247
Score: 176.00 Matches: 47
Percent Similarity: 51.09% Conservative: 23
Best Local Similarity: 34.31% Mismatches: 41
Query Match: 16.76% Indels: 26
DB: Gaps: 5

```

US-09-867-753-1 (1-577) x US-09-129-888-2 (1-247)

```

QY 122 GAAGGC---CATGTTGGCCCAAGAGCTCCAGGCTCATGCTAATATGAACCTTGAGGC 178
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 99 GlnGlyArgHisAlaGlyAspGlyAla----- 107
QY 179 GGTGTGAACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 238
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 108 -----SerSerSerGlnAspAspSerIleuIleuIleuIleuIleuIleuIleu 121
QY 239 GGA---AACCAAGAGCTGCGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 295
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 122 AenIleAspGlnGlnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 136
QY 226 GCCATGAGAGGTCGCGAGCGCGAGAAC-----ATGACACCAAGCAAGCAAGCAAGCAAG 349
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 137 SerIleArgAsnProHisValIleuAsnArgIleuAlaGlnIleuArgIleuArgIleuArg 156
QY 350 TTCACGCTGTTCAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGT 409
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 157 PheThrHisSerGlnIleuHisAspIleuGlnIleuGlnIleuGlnIleuGlnIleuGln 176
QY 410 GTGCCCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 469
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 177 LeuArgAlaArgArgAspLeuAlaArgTrpMetGlyValAspGlnIleuGlnIleuGln 196
QY 470 TGGTTTAAGATTAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 520
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 197 TrpPheArgMetArgArgIleuPheGlnIleuArgAsnArgArgValIleuMet 213

```

RESULT 8

US-09-105-470B-4

; Sequence 4, Application US/09105470B

; Patent No. 6617129

; GENERAL INFORMATION:

; APPLICANT: Carter, Kenneth C.

; He, Wei-Mu

; TITLE OF INVENTION: Human NK-3 Related Prostate Specific

; Gene-1

; NUMBER OF SEQUENCES: 30

; CORRESPONDENCE ADDRESS: KESSLER, GOLDSTEIN & FOX P.L.L.C.

; ADDRESS: 1100 NEW YORK AVE., NW, SUITE 600

; CITY: WASHINGTON

; STATE: DC

; COUNTRY: USA

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/105,470B

; FILING DATE: 26-Jun-1998

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/051,080

; FILING DATE: 27-JUN-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: STEFFE, ERIC K.

; REGISTRATION NUMBER: 36,688

; REFERENCE/DOCKET NUMBER: 1488,0790001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 301-309-8504

; TELEFAX: 301-309-8439

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 234 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 4:

US-09-105-470B-4

Alignment Scores:

Pred. No.: 1,49e-08 Length: 234

Score: 174.00 Matches: 60

Percent Similarity: 39.81% Conservative: 22

Best Local Similarity: 29.13% Mismatches: 57

Query Match: 16.57% Indels: 67

DB: Gaps: 8

US-09-867-753-1 (1-577) x US-09-105-470B-4 (1-234)

```

QY 58 CTGCTGAGTGTATCCAGGTAAT-----AAT 84
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4 ValProGlnProArgProGlyGlnAlaValAspGlnGlnAlaProProThrProSer 23
QY 85 AAGCCCAACCTCA-----GCTGGG 105
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24 LysProLeuThrSerPheLeuIleGlnAspIleLeuAspGlnGlnAlaArgGlnGln 43
QY 106 GGCAGCATCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 165
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 44 GlyArgThrSerSerGlnArgGlnIleAspPro----- 54
QY 166 GAACCTGAGGCGGTGTGAACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 225
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55 -----GlnProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGln 67
QY 226 CGAGGCGCGGTGGAACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 285

```

```

Db      67 rargalaglyalaglnasbplneuserThrglyProargAlaAlaProgluGlun1 87
Qy      286 GGCCCGAGGGGCCATGAGGGGTCCGAG-----CCCA 318
Db      87 aglnrleuAlaGluthrGlInProgluInrghlsleuglyserTyLeuleuaspsergl 107
Qy      319 GAAC-----ATGCAGCCAGCAACTCGG-- 340
Db      107 uanenthserGlyAlaLeuProargLeuProglInthrProlyeglInProglInysargse 127
Qy      341 -CGCAGAAAGTTACCGCTGTTCAGAGCTGAGAGCTTGAAAGTCTTTCCGACACACTCA 399
Db      127 rargalaaAlaPheSerhlsThrglnvalleugluargLysPheSerhlsGlnly 147
Qy      400 ATACCTGTATGTGCCCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 459
Db      147 sTyLeuSerAlaAlaProgluInrghlsleuAlaLysAsnleuLysleuThrglnThrgl 167
Qy      460 AGTCCGGGTTTGTTCAGATTAAGATTAAGGCGCAGATGTAGCGCATCAGAGAAATTAAT 519
Db      167 nvallyslletrPheGlnAsnArgArgTrlyrThrlYsArgLysglIn----- 183
Qy      520 GCTGCCAATGAACTA 535
Db      184 -LeuSerSerGlnleu 188

```

RESULT 9

```

US-09-976-594-443
; Sequence 443, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 443
; LENGTH: 185
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 168861CD1
US-09-976-594-443

```

Alignment Scores:

```

Pred. No.: 1 75e-08 Length: 185
Score: 173.00 Matches: 53
Percent Similarity: 44.83% Conservative: 25
Best Local Similarity: 30.46% Mismatches: 53
Query Match: 16.48% Indels: 43
DB: 4 Gaps: 7

```

```

US-09-867-753-1 (1-577) x US-09-976-594-443 (1-185)
Qy      76 GGTAAATTAAGCCACACCTGAGTGGGGGAGATCAAGCGC----- 120
Db      12 GlyGlnAsnLysProSerThr-----CysSerPheSerlleGluArglleuGlyLeuasp 30
Qy      121 ---AGAGGCGCATGTTGGCCA-----AGAGCTCCAGGCTTATGGGTAATTAATGAACC 171
Db      31 GlnLysLysAspCysValProleuMetLysProHisArgPro-TripAlaAspThrCysse 50
Qy      172 TGAAGGCGGCGGTGAACACGAGAAAGGCGATGAACGGGATGCGCGGATG-----AT 222
Db      50 rSerSerGly-----LysAspGlyAsnleuLysleuHisVal 62

```

```

Qy      223 CCCCGAGGGCGCGGTGAAACACGAGCCTCGGACGAGCCGCCCGGAGAGA 282
Db      62 lProAsnProProSerGlylIleSerPheProSerValValaPheIlePrometProgluGl 82
Qy      283 GCCCGCCGAGGGGCCCATGAGGGGTCCGAGCCCGAGAAC----- 322
Db      82 uArgAlaSer-----LysTyrglnAsnTyPheSerAlaSerGlnur 96
Qy      323 -----ATGCAGCCACGAACTCGGCGCAGCAAGTTTCAC 354
Db      96 glenSerleuLysArgGlnleuSerTrpTyTrgGlyArgArgProArgThrlAlaPheTh 116
Qy      355 GCTGTTCAGGTGAGAGAGCTGGAAGTCTTTCCGACACACTCAATACCTGATGTGCC 414
Db      116 rGlnAsnGlnlleGluValleuGlnAsnValPheArgValaAsnCyTyProGlylLeas 136
Qy      415 CACAGAGAGGAACTTGCCGAAACTTACGTGTGACTGAGAGCAAGTCGGGTTTGGTT 474
Db      136 pIleArgGlnAspLeuAlaGlnLysleuAsnleuGlnAspArglleGlnlletrPhe 156
Qy      475 TAAGATTAAGGCGCAGATGTAGCGCATCAGAGAGA 514
Db      156 eGlnAsnArgArgAlaLysleuLysArgSerSerhlsArgGln 169

```

RESULT 10

```

US-09-976-594-945
; Sequence 945, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 945
; LENGTH: 289
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 1505790CD1
US-09-976-594-945

```

Alignment Scores:

```

Pred. No.: 2 7e-08 Length: 289
Score: 171.50 Matches: 55
Percent Similarity: 42.11% Conservative: 33
Best Local Similarity: 26.32% Mismatches: 74
Query Match: 16.33% Indels: 47
DB: 4 Gaps: 8

```

```

US-09-867-753-1 (1-577) x US-09-976-594-945 (1-289)
Qy      2 CGAATCAAGGCGCTCCACCATGCGGCTGTGCTGCACAGACCGGTCTTACTATGC 61
Db      30 ProSerGlnGlnSerProThrleuProGlnSerSerAlaThrsPheSerPtyrTy--- 48
Qy      62 CTGAGTGATACAGGTAAATAAGCCCGCACACCTCATGCTGGGGGACACATCAAGCGCA 121
Db      49 -----SerProThr-----GlyGlyAlaPro 55
Qy      122 GAAGGCGATGTGGCCAGAGCTCCAGGCTTCAGGAGTAATTAATGAACCTGAAGGCGGT 181
Db      56 HisGlyTyTrCysSerProThrSerAlaSerTyGlyLysAlaLeuAsnProTyTrGlnTy 75
Qy      182 GTGAACACAGAGAGCGGATGAACCGCGATGCGGCGCATATCCCGAG----- 229
Db      76 GlnTyTrHis-----GlyValAsnGlnLysAlaGlySerTyProAlaLysAlaTyAla 93

```

```

QY 220 -----GGCGGCGGTGAAACCGAGGCTCGG 256
DB 94 AspyrSeryrAlaSerSerTyrrHisGlnrYrgIyGlyAlaTyrrAsnArgValProSer 113
QY 257 CAGCAGCCGACCCCGCCGAG-----GAGCCGCGCCAGCGCGCCATGAGAGGCTCGG 310
DB 114 AlatrHsngInProGlnrYsgIValThrGluProGlnrValArgMetValAsnGlyLys 133
QY 311 CAGCCCGAGAAATGATGACGACCACTCGGCGCAAGATTTCACGCTGTTCAGAGTGGAG 370
DB 134 ProLysrValArLysrProArGthrIleTyrrSerSerPhe-----GlnLeuAla 150
QY 371 GAGCTGAAAGTGTTCGACACACTCAATACCCGTGATGCCCGACAGAGGAACTT 430
DB 151 AlaleuGlnArArgPheGlnrYsThrGlnrYrLeuAlaLeuProGlnrArgAlaGluLeu 170
QY 431 GCGAAACTTGAAGTGTGACTGAAGACAAAGTCCGGTGTGTTAAGATTAAGAGGCC 490
DB 171 AlaAlaSerLeuGlyLeuThrGlnrYrValrYsIleTrpPheGlnrAsnLysArgSer 190
QY 491 AGATGTAGCCGACATCAGAGAGAAATTAATGCTCCCAATGAACTA----- 535
DB 191 LysIleLysLys-----IleMetLysAsnGlyGluMetProGlnrHisSer 206
QY 536 -----CGTGTGACCCAGACGACTGT 556
DB 207 ProSerSerSerAspProMetAlaCys 215

```

RESULT 11

US-08-957-351-26
Sequence 26, Application US/08957351
Patent No. 6306586

GENERAL INFORMATION:

APPLICANT: Semina, Elena
APPLICANT: Murray, Jeffrey C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HONG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109-2170

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/957,351
FILING DATE: 24-OCT-1997
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: UIA-024.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:
LENGTH: 271 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-957-351-26

Alignment Scores:

Pred. No.: 4.13e-08 Length: 271

Score: 169.50 Matches: 38
Percent Similarity: 54.29% Conservative: 19
Best Local Similarity: 36.19% Mismatches: 29
Query Match: 16.14% Indels: 19
DB: 4 Gaps: 3

US-09-867-753-1 (1-577) x US-08-957-351-26 (1-271)

```

QY 197 GGCATGAACCCGATGGGCGGATATCCCGAGGCGCGGTGAAACCGAGGCTCGG 256
DB 14 GtLysGlnrYsAspLysGlyGln-----GlnGlyLysAsnGluAsp----- 27
QY 257 CAGCAGCCGACCCCGCCGAGAGCGCGCCAGCGCGCCATGAGAGGCTCGGAGGCC 316
DB 28 -----ValGlyAlaGluAspProSerLys 35
QY 317 GAGAAATGACGACCCAGCAACTCGGCGCAAGATTTCAGGCTTTCAGAGTGAAGAGTCTG 376
DB 36 LysLysArgGln---ArgArgGlnrYrHisPheThrSerGlnGlnrLeuGlnrLeu 54
QY 377 GAAAGTGTTCGACACACTCAATACCCGTGATGCCCGACAGAGGAACTTCCGAA 436
DB 55 GluAlaThrPheGlnrArgAsnArgTyrrProAspMetSerThrArgGluGlnrIleAlaVal 74
QY 437 AACTTGAAGTGTGACTGAAGACAAAGTCCGGTGTGTTAAGATTAAGAGGCCAGATGT 496
DB 75 TrpThrAsnLeuThrGlnrAlaArgValrYrPheLysAsnArgArgAlaLysTrp 94
QY 497 AGCCGACATCAGAGA 511
DB 95 ArgLysArgGluArg 99

```

RESULT 12

US-09-636-215-842
Sequence 842, Application US/09636215
Patent No. 6620922

GENERAL INFORMATION:

APPLICANT: Xu, Jianshun
APPLICANT: Dillon, Devin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuqul
APPLICANT: Henderson, Robert A.
APPLICANT: Kalos, Michael D.
APPLICANT: Ranger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darrick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yaser A.W.
TITLE OF INVENTION: Hepier, William
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.4271C17
CURRENT APPLICATION NUMBER: US/09/636,215
NUMBER OF SEQ ID NOS: 852
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 842
LENGTH: 241
TYPE: PRT
ORGANISM: Homo sapiens
US-09-636-215-842

Alignment Scores:

Pred. No.: 4.48e-08 Length: 241
Score: 169.00 Matches: 60
Percent Similarity: 39.52% Conservative: 23
Best Local Similarity: 28.57% Mismatches: 60
Query Match: 16.10% Indels: 67

```

DB: 4 Gaps: 8
US-09-867-753-1 (1-577) x US-09-636-215-842 (1-241)
QY 46 CACCGGTCTACTGCTGAGTGTATACCAAGTAA----- 81
DB 7 HSHSLLeuArgValProGluProArgProGluAlaLysAlaGluGlyAlaAlaPro 26
QY 82 -----AATAGCCCAACCTCA----- 99
DB 27 ProThrProSerLysProLeuThrSerPheLeuIleGlnAspIleLeuArgAspGlyAla 46
QY 100 -----GCTGGGGGAGCATCAAGCCGAAGGCCATGTTGGCCAAAGAGCTCCAGCCCT 153
DB 47 GlnArgGlnGlyGlyArgThrSerSerGlnArgGlnArgAspPro----- 61
QY 154 CATGGTAAATATGAACCTGAGGGCGGTGTGAACCAAGAACGAGATGAACCGCATGG 213
DB 62 -----GlnProGluProGluProGluPro-GluGlu 71
QY 214 CGCATGATCCCGAGGCGCGGTGGAACCAAGAGCTCCGAGAGCCGAGCCGCC 273
DB 71 yGlyArg---SerArgAlaGlyAlaGlnAsnAspGlnLeuSerThrGlyProArgAlaAl 90
QY 274 GCCGAGAGAGCCGCGCCAGCGCCATGAGGGTCCGAG----- 313
DB 90 aProGluGluAlaGluThrLeuAlaGluThrGluProGluArgHisLeuGlySerTyrIle 110
QY 314 -----CCCGAGAAC-----ATGAGAGCC 330
DB 110 uLeuAspSerGluAsnThrSerGlyAlaLeuProArgLeuProGlnThrProLysGlnPr 130
QY 331 ACGAAGTCCG---CGCAGCAAGTTCAAGCTGTTCAGGTGAGAGAGCTGGAAGTGT 387
DB 130 oGlnLysArgSerArgAlaAlaPheSerHisThrGlnValIleGlnLeuGlnArgLysPh 150
QY 388 CCGACACTCAATACCTGATGTGCCCAAGAGAGGAACTTGGCGAAACTTAGGTGT 447
DB 150 eSerHisGlnLysTyrLeuSerAlaProGluArgAlaHisLeuAlaLysAsnLeuLysIle 170
QY 448 GACTGAAGCAAAAGTCGGGTTGGTTTAAGATAAAGGCGCAGATGAAGGAGCATCA 507
DB 170 uThrGluThrGlnValLysIleTyrPheGlnAsnArgArgTyrLysThrLysArgLysG 190
QY 508 GAGAGAAATTAATGCTCGCAATGAACCTA 535
DB 190 n-----LeuSerSerGlnLeu 195

RESULT 13
US-09-685-166A-842
; Sequence 842, Application US/09685166A
; Patent No. 6630305
; GENERAL INFORMATION:
; APPLICANT: Xu, Jianshun
; APPLICANT: Dillon, Devin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Heplet, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C21

```

```

; CURRENT APPLICATION NUMBER: US/09/685,166A
; CURRENT FILING DATE: 2000-10-10
; NUMBER OF SEQ ID NOS: 898
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 842
; LENGTH: 241
; TYPE: PRN
; ORGANISM: Homo sapiens
US-09-685-166A-842

Alignment Scores:
Pred. No.: 4,48e-08 Length: 241
Score: 169.00 Matches: 60
Percent Similarity: 39.52% Conservative: 23
Best Local Similarity: 28.57% Mismatches: 60
Query Match: 16.10% Indels: 67
DB: 4 Gaps: 8

US-09-867-753-1 (1-577) x US-09-685-166A-842 (1-241)
QY 46 CACCGGTCTACTGCTGAGTGTATACCAAGTAA----- 81
DB 7 HSHSLLeuArgValProGluProArgProGluAlaLysAlaGluGlyAlaAlaPro 26
QY 82 -----AATAGCCCAACCTCA----- 99
DB 27 ProThrProSerLysProLeuThrSerPheLeuIleGlnAspIleLeuArgAspGlyAla 46
QY 100 -----GCTGGGGGAGCATCAAGCCGAAGGCCATGTTGGCCAAAGAGCTCCAGCCCT 153
DB 47 GlnArgGlnGlyGlyArgThrSerSerGlnArgGlnArgAspPro----- 61
QY 154 CATGGTAAATATGAACCTGAGGGCGGTGTGAACCAAGAACGAGATGAACCGCATGG 213
DB 62 -----GlnProGluProGluProGluPro-GluGlu 71
QY 214 CGCATGATCCCGAGGCGCGGTGGAACCAAGAGCTCCGAGAGCCGAGCCGCC 273
DB 71 yGlyArg---SerArgAlaGlyAlaGlnAsnAspGlnLeuSerThrGlyProArgAlaAl 90
QY 274 GCCGAGAGAGCCGCGCCAGCGCCATGAGGGTCCGAG----- 313
DB 90 aProGluGluAlaGluThrLeuAlaGluThrGluProGluArgHisLeuGlySerTyrIle 110
QY 314 -----CCCGAGAAC-----ATGAGAGCC 330
DB 110 uLeuAspSerGluAsnThrSerGlyAlaLeuProArgLeuProGlnThrProLysGlnPr 130
QY 331 ACGAAGTCCG---CGCAGCAAGTTCAAGCTGTTCAGGTGAGAGAGCTGGAAGTGT 387
DB 130 oGlnLysArgSerArgAlaAlaPheSerHisThrGlnValIleGlnLeuGlnArgLysPh 150
QY 388 CCGACACTCAATACCTGATGTGCCCAAGAGAGGAACTTGGCGAAACTTAGGTGT 447
DB 150 eSerHisGlnLysTyrLeuSerAlaProGluArgAlaHisLeuAlaLysAsnLeuLysIle 170
QY 448 GACTGAAGCAAAAGTCGGGTTGGTTTAAGATAAAGGCGCAGATGAAGGAGCATCA 507
DB 170 uThrGluThrGlnValLysIleTyrPheGlnAsnArgArgTyrLysThrLysArgLysG 190
QY 508 GAGAGAAATTAATGCTCGCAATGAACCTA 535
DB 190 n-----LeuSerSerGlnLeu 195

RESULT 14
US-08-754-477A-18
; Sequence 18, Application US/08754477A
; Patent No. 6518411
; GENERAL INFORMATION:
; APPLICANT: Murray, Jeffrey
; APPLICANT: Semina, Elena
; TITLE OF INVENTION: RIBS COMPOSITIONS AND THERAPEUTIC
; TITLE OF INVENTION: AND DIAGNOSTIC USES THEREFOR

```

NUMBER OF SEQUENCES: 139
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109-2170
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/754,477A
FILING DATE: 22-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: UIA-022.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 60 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-08-754-477A-18

Alignment Scores:
Pred. No.: 5.04e-08 Length: 60
Score: 167.00 Matches: 31
Percent Similarity: 71.19% Conservative: 11
Best Local Similarity: 52.54% Mismatches: 17
Query Match: 15.90% Indels: 0
Gaps: 0

US-09-867-753-1 (1-577) x US-08-754-477A-18 (1-60)
QY 332 CGAAGTCGGCGCAGAGTTTCAGCTGTTCAGGTGAGAGCTGGAAGTGTTCGCA 391
DB 2 ArgAGGTGTATGTThrPheThrsrPheGlnLeuGlnLeuGlnLysAlaPheSer 21
QY 392 CACACTCAATACCTGATGTGCCACAGAGAGGAACTTGCGAAACTTAGGTGACT 451
DB 22 ArgThrHisTyrProAspValPheThrArgGlnGlnLeuAlaMetLysIleGlyLeuThr 41
QY 452 GAAGCAAAAGTCGGGTTTGGTTTAAGATTAAGGCGCAGATGTAGCGGACATCAG 508
DB 42 GtualaAArgIleGlnValTTPheGlnAsnArgAlaLysTTPArgLysGlnGln 60

RESULT 15
US-08-754-477A-19
Sequence 19, Application US/08754477A
Patent No. 6518411
GENERAL INFORMATION:
APPLICANT: Murray, Jeffrey
APPLICANT: Semina, Elena
TITLE OF INVENTION: RING COMPOSITIONS AND THERAPEUTIC
TITLE OF INVENTION: AND DIAGNOSTIC USES THEREFOR
NUMBER OF SEQUENCES: 139
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109-2170
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/754,477A
FILING DATE: 22-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: UIA-022.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 60 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-08-754-477A-19

Alignment Scores:
Pred. No.: 5.04e-08 Length: 60
Score: 167.00 Matches: 31
Percent Similarity: 71.19% Conservative: 11
Best Local Similarity: 52.54% Mismatches: 17
Query Match: 15.90% Indels: 0
Gaps: 0

US-09-867-753-1 (1-577) x US-08-754-477A-19 (1-60)
QY 332 CGAAGTCGGCGCAGAGTTTCAGCTGTTCAGGTGAGAGCTGGAAGTGTTCGCA 391
DB 2 ArgAGGTGTATGTThrPheThrsrPheGlnLeuGlnLeuGlnLysAlaPheSer 21
QY 392 CACACTCAATACCTGATGTGCCACAGAGAGGAACTTGCGAAACTTAGGTGACT 451
DB 22 ArgThrHisTyrProAspValPheThrArgGlnGlnLeuAlaMetLysIleGlyLeuThr 41
QY 452 GAAGCAAAAGTCGGGTTTGGTTTAAGATTAAGGCGCAGATGTAGCGGACATCAG 508
DB 42 GtualaAArgIleGlnValTTPheGlnAsnArgAlaLysTTPArgLysGlnGln 60

Search completed: June 16, 2004, 19:04:07
Job time: 21.5 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 16, 2004, 18:56:12 ; Search time 46 Seconds

(without alignments)
7915.391 Million cell updates/sec

Title: US-09-867-753-1

Perfect score: 1050

Sequence: 1 tccacacacagcgcctccag.....tctacatcgctgtgactag 577

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+np.model -DBV=rlp
-Q=/cgn2.1/USPTO.spool.p/US09867753/runac.16062004.163635.27793/app_query.fasta.1.775
-DB=SPTRMBL.25 -QFWT=faetan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -ALIGN=15 -MODE=LOCAL
-OUTPUT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09867753@CEN.1.1.86@runac.16062004.163635.27793 -NCPV=6 -ICPV=3
-NO MAP -LANG=ENGLISH -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTRMBL.25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp Vertebrate:*
14: sp Unclassified:*
15: sp_rv1rns:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARYS

Result No.	Score	Query Match	Length	DB ID	Description
1	298.5	28.4	406	Q8N693	Q8N693 homo sapien

2	298.5	28.4	406	4	Q7Z6K7	Q7Z6K7 homo sapien
3	212.5	20.2	227	11	Q9E0M5	Q9E0M5 mus musculus
4	212	20.2	286	5	Q8T6T5	Q8T6T5 branchiostro
5	202.5	19.3	317	13	Q8A1U4	Q8A1U4 gallus gall
6	200.5	19.1	391	5	Q8T8C1	Q8T8C1 gryllus bim
7	200	19.0	248	13	Q12952	Q12952 cynops pyrr
8	197	18.8	382	11	Q88933	Q88933 mus musculus
9	197	18.8	387	11	Q922U3	Q922U3 mus musculus
10	196	18.7	314	11	Q54817	Q54817 mus musculus
11	196	18.7	408	5	Q9VPE1	Q9VPE1 drosophila
12	192	18.3	328	13	Q9W7M5	Q9W7M5 brachydanio
13	190	18.1	227	11	Q8BPD6	Q8BPD6 mus musculus
14	190	18.1	429	5	Q7Z0W2	Q7Z0W2 lycichinus
15	190	18.1	640	5	Q8T0M4	Q8T0M4 dirosophila
16	189.5	18.0	325	11	Q80YJ6	Q80YJ6 mus musculus
17	188.5	18.0	371	5	Q61282	Q61282 hemisceroc
18	188	17.9	227	11	Q70Z38	Q70Z38 mus musculus
19	187	17.8	257	4	Q86YR1	Q86YR1 homo sapien
20	185.5	17.7	430	5	Q7Z0W3	Q7Z0W3 strongyloce
21	184.5	17.6	226	6	Q9GLL9	Q9GLL9 sus scrofa
22	184.5	17.6	318	5	Q8EB70	Q8EB70 drosophila
23	183.5	17.5	287	5	Q8WNG3	Q8WNG3 patella val
24	183	17.4	299	13	Q7SZN8	Q7SZN8 brachydanio
25	183	17.4	826	5	Q9B130	Q9B130 ciona intes
26	182.5	17.4	185	13	P79857	P79857 pleurodeles
27	182	17.3	374	5	Q81EX1	Q81EX1 ciona intes
28	181.5	17.3	240	11	Q8R4I3	Q8R4I3 mus musculus
29	180	17.1	185	11	Q8C477	Q8C477 mus musculus
30	180	17.1	299	6	Q8SQ03	Q8SQ03 canis famil
31	180	17.1	299	11	Q9WTQ9	Q9WTQ9 rattus norv
32	180	17.1	299	11	Q9JLRE	Q9JLRE rattus norv
33	179.5	17.1	362	5	Q21836	Q21836 caenorhabdi
34	179	17.0	301	5	Q46170	Q46170 tribolium c
35	177.5	16.9	276	13	Q98TG7	Q98TG7 gallus gall
36	177	16.9	387	13	Q93582	Q93582 gallus gall
37	176.5	16.8	394	5	Q9XYQ3	Q9XYQ3 anopheles g
38	176.5	16.8	503	11	Q9ES16	Q9ES16 mus musculus
39	176	16.8	214	13	Q8UVD3	Q8UVD3 gallus gall
40	176	16.8	284	4	Q81VZ2	Q81VZ2 homo sapien
41	176	16.8	370	5	Q25411	Q25411 lineus sang
42	176	16.8	403	4	Q86UQ3	Q86UQ3 homo sapien
43	176	16.8	407	4	Q86UQ2	Q86UQ2 homo sapien
44	176	16.8	479	11	Q9CX16	Q9CX16 mus musculus
45	176	16.8	479	11	Q8BRE7	Q8BRE7 mus musculus

ALIGNMENTS

RESULT 1
ID Q8N693 PRELIMINARY; PRT; 406 AA.
AC Q8N693;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE ESX1L.
GN ESX1L.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pooled germ cell tumors;
RX MEDLINE=2169184; PubMed=11374906;
RA Fohn L.E., Behringer R.R.;
RT "ESX1L, a novel X chromosome-linked human homeobox gene expressed in the placenta and testis";
RL Genomics 74:105-108(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Pooled germ cell tumors;
RA Fohn L.E., Behringer R.R.;

Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

RL EMBL: AY114148; AAM62141.1; -
 DR GeneW: HGNC:14865; ESX1L.
 DR GO: GO:0005634; C:nucleus; IEA.
 DR GO: GO:0003700; E:transcription factor activity; IEA.
 DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro: IPR001356; Homeobox.
 DR InterPro: IPR007104; Paired homeo.
 DR InterPro: IPR000327; POU domain.
 DR Pfam: PF00046; homeobox; 1.
 DR PRINTS: PRO0028; POUDOMAIN.
 DR ProDom: PD00010; Homeobox; 1.
 DR SMART: SM00389; HOX; 1.
 DR PROSITE: PS00027; HOMEBOX_1; 1.
 DR PROSITE: PS00071; HOMEBOX_2; 1.
 SO SEQUENCE 406 AA; 44296 MW; 701B3986F1148FA CRC64;

Alignment Scores:

Pred. No.:	1.16e-18	Length:	406
Score:	298.50	Matches:	71
Percent Similarity:	49.21%	Conservative:	22
Best Local Similarity:	37.57%	Mismatches:	39
Query Match:	28.43%	Indels:	57
DB:	4	Gaps:	6

US-09-867-753-1 (1-577) x Q8N693 (1-406)

QY 170 CCTGAGGCGGCTGTGAACCGCATGACCGCGATGCGGCGATGATCCGCCC--- 226
 DB 51 ProglutrylthrlglutlalaagluamenaValglThrlglutlyser---ValProser 69
 QY 227 -----GAGGCGCGGCTGTGAACCGAGAGCGCTGTGAGCGAGCGCGAGCC 271
 DB 70 AspaRglnaAspaRglnuglyglYglYglYHsIglProglInglInglInglInglPro 89
 QY 272 CGC----- 274
 DB 90 ProleuThrlYbProglInglInglInglInglInglProProleuLeuIngluLeuYsgInglu 109
 QY 275 CCGGAGAGCGCGCGCCGCGAGCGCGCGAGGTCGCGAGCC----- 316
 DB 110 GlnGlIngluProProglInthrThrlValgluglyProglInProalaglInglYProglInthr 129
 QY 317 ---GAGAACATGACGCA-----CGAACTCGCGCGAGAGTTTCACGCTGTG 361
 DB 130 AlaGluglyProglInProProgluArglyArgArgArgRghrlaPhethrInglInphe 149
 QY 362 CAGGTGAGAGAGCTGGAAGTGTTCGACACACTCAATCACTGATGTGCGCCACAAGA 421
 DB 150 GlnLeuGlnIngluLeuInsnPheAspGlnSerGlnTrpAspValValAlaArg 169
 QY 422 AGGGAAGTCCGAAAGTGTGACTGAGAGCAAGAACTGCGGCTTGGTTTAAAGAT 481
 DB 170 GluArgLeuAlaAlaArgLeuLeuThrlGlnAspArgValGlnValTrpPheGlnAsn 189
 QY 482 AAAAGGCGCATGTGTGCGACATCAGAGAAATTATGCTCGCAATGAACTACGCTGTG 541
 DB 190 ArgArgAlaArgTrpArgAspGlnArgValLeuLeuLeuArgsnThrAlaThrAla 209
 QY 541 ----- 541
 DB 210 AspleuAlaHisProleuAspMetPheLeuGlyGlyAlaTyTrpAlaAlaProAlaLeu 229
 QY 542 GACCCAGAGAGCTGTGTATCATCGTC 568
 DB 230 AspProAlaLeuCYsValHisLeuVal 238

RESULT 2
 Q726K7 PRELIMINARY; PRT; 406 AA.
 AC Q726K7;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Extremebryonic, spermatogenesis, homeobox 1-like.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Scheifter C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaney S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smallus D.E., Scherch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Strausberg R.;
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC053599; AAH53599.1; -
 KW DNA-binding; Homeobox; Nuclear protein.
 SQ SEQUENCE 406 AA; 44311 MW; D8B9C18FB9235B9 CRC64;

Alignment Scores:

Pred. No.:	1.16e-18	Length:	406
Score:	298.50	Matches:	71
Percent Similarity:	49.21%	Conservative:	22
Best Local Similarity:	37.57%	Mismatches:	39
Query Match:	28.43%	Indels:	57
DB:	4	Gaps:	6

US-09-867-753-1 (1-577) x Q726K7 (1-406)

QY 170 CCTGAGGCGGCTGTGAACCGCATGACCGCGATGCGGCGATGATCCGCCC--- 226
 DB 51 ProglutrylthrlglutlalaagluamenaValglThrlglutlyser---ValProser 69
 QY 227 -----GAGGCGCGGCTGTGAACCGAGAGCGCTGTGAGCGCGAGCC 271
 DB 70 AspaRglnaAspaRglnuglyglYglYglYHsIglProglInglInglInglInglPro 89
 QY 272 CGC----- 274
 DB 90 ProleuThrlYbProglInglInglInglInglInglProProleuLeuIngluLeuYsgInglu 109
 QY 275 CCGGAGAGCGCGCGCCGCGAGCGCGAGGTCGCGAGCC----- 316
 DB 110 GlnGlIngluProProglInthrThrlValgluglyProglInProalaglInglYProglInthr 129
 QY 317 ---GAGAACATGACGCA-----CGAACTCGCGCGAGAGTTTCACGCTGTG 361
 DB 130 AlaGluglyProglInProProgluArglyArgArgArgRghrlaPhethrInglInphe 149
 QY 362 CAGGTGAGAGAGCTGGAAGTGTTCGACACACTCAATCACTGATGTGCGCCACAAGA 421
 DB 150 GlnLeuGlnIngluLeuInsnPheAspGlnSerGlnTrpAspValValAlaArg 169

```

QY 422 AGGAACTTGGCCGAAACTTAGATGCTAGAGCAAGAGTGGGTTGTTTAAGAT 481
DB 170 GtUaGtLeuAdAlaArgLeuAmLeuThrGluAspArgValGlnValTrpPheGlnAsn 189
QY 482 AAAAGGCGGCAAGTGTATGAGCATGACAGAGATTAATCTCGCCCAATGACCTGCT 541
DB 190 ArgGdAlaLysTrpLysArgAsnGlnArgValLeuMetLeuArgSerThrAlaThrAla 209
QY 541 ----- 541
DB 210 AepLeuAlaHlePheLeuAepMetPheLeuGlyAlaTyrTyrAlaAlaProAlaLeu 229
QY 542 GACCCAGACAGCTGTGTACATCGTC 568
DB 230 AepProAlaLeuCyValHisLeuVal 238

RESULT 3
Q9EOM5 PRELIMINARY; PRT; 227 AA.
ID Q9EOM5:
AC Q9EOM5:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Homeobox protein GBPOX.
GN PSX2 OR 1600026001RIK OR GBPOX.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Takasaki N., McIsaac R., Dean J.;
RT "Gpox, a novel homeobox gene preferentially expressed in female germ
RT cells at the onset of sexual dimorphism in mice.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF201698; AAG36768.1; -.
DR MGD; MGI:1890128; Pox2.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; P:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
SQ SEQUENCE 227 AA; 25112 MW; D39CF6795AF682DC CRC64;

Alignment Scores:
Pred. No.: 7 596-11 Length: 227
Score: 212.50 Matches: 61
Percent Similarity: 45.61% Conservative: 17
Best Local Similarity: 35.67% Mismatches: 60
Query Match: 20.24% Indels: 33
DB: 11 Gaps: 6

US-09-867-753-1 (1-577) x Q9EOM5 (1-227)
QY 104 GGGGAGCATCAAGCCAGGCGCATGTTGGCCAGAGCTCCAGGCTCATGGTAAT 163
DB 54 GlyGlyLeuAepGlnGlyGluGlnGlyAlaGlnGlyValAlaGlyGlyGluGlnAlaGln 73
QY 164 ATGACACCT-----GAGGCGGTGTGACACCAAGAAAGCGCATG 202
DB 74 GluGluProAlaProLeuSerProAlaGlnGluAlaThrGlyGlyGluGluGluGlu 93
QY 203 AACCGGATGGGCGCATG----- 220
DB 94 AsnLysGlnGlyGluMetGluGlyAlaGlyAlaGlyAlaSerGlyProGluAsp 113
QY 221 -----ATCCCGAGGCGCGGTGGAAC-----CAGAGGCTCGAGAGCCGCGAG 268
DB 114 AspAsnIleGlnGlnGluGluGlyGluGluGlnIleAepGlnGlnProProGlnGlnGluAla 133

```

```

QY 269 CCCCCCGAGAGAGCCGCGCCAGGCGGCGATGAGAGGTCCGACCCGGAAC----- 322
DB 134 AlaIleProGlu-----GlyMetArgAsnProGlnAlaGlyAsnTyrLeu 148
QY 322 ATGACAGCAAGAACTCGGCGCAGAGTTTCAAGCTGTGGCAGAGAGCTGGAAGT 382
DB 149 AlaHisGlnArgThrArgArgThrArgPheThrHisSerGlnLeuArgSerGluArg 168
QY 383 GTTTTCCGACAGATCAATACCTGATGTGCCACAGAGGAAGAACTTCCGAAACTTA 442
DB 169 LeuPheGlnGluAsnArgPheProSerLeuArgValArgArgAspLeuAlaArgTyrMet 188
QY 443 GGTGTGAGTGAAGACAAAGTGGGTTGGTTTAAGATTAAGAGGCGCAGATGTAGCGCA 502
DB 189 GlyValAspGluSerAspValGlnGluTrpPheLysMetArgAlaLeuPheArgArg 208
QY 503 CATCAGAGAAATTAATGCTCGCCCAATGAATA 535
DB 209 HisSerArg--LeuMetMetPheCysGluLeu 218

RESULT 4
Q8T615 PRELIMINARY; PRT; 286 AA.
ID Q8T615:
AC Q8T615:
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Paired-like homeodomain transcription factor Shox.
GN SHOX.
OS Branchiostoma floridae (Florida lancelet) (Amphioxus).
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC Branchiostoma.
OX NCBI_TaxId=7739;
RN [1]
RP SEQUENCE FROM N.A.
RA Jackman W.R., Jr., Kimmel C.B.;
RT "Coincident iterated gene expression in the amphioxus neural tube.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; AF465939; AAL83210.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; P:transcription factor activity; IEA.
DR GO; GO:0007275; P:development; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR003654; Homeo_OAR.
DR InterPro; IPR00047; HTH_lambdarepressr.
DR InterPro; IPR007104; Paired_homeo.
DR Pfam; PF00046; homeobox; 1.
DR Pfam; PF03826; OAR; 1.
DR PRINTS; PR00031; HTHREPRESSR.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
DR PROSITE; PS50803; OAR; 1.
DR PROSITE; PS50803; OAR; 1.
KW Homeobox; DNA-binding; Nuclear protein.
SQ SEQUENCE 286 AA; 31551 MW; 6BF50B62D92958C1 CRC64;

Alignment Scores:
Pred. No.: 8 66-11 Length: 286
Score: 212.00 Matches: 52
Percent Similarity: 48.10% Conservative: 24
Best Local Similarity: 32.91% Mismatches: 48
Query Match: 20.19% Indels: 34
DB: 5 Gaps: 5

US-09-867-753-1 (1-577) x Q8T615 (1-286)
QY 104 GGGGAGCATCAAGCGCAGAGGCGCAT-----GTTGGCCAGAGCTCCAGGCTCATG 157
DB 46 GlyAlaGluSerAspValAspGlyAspAspLeuCySProValGlySerProGlyLeu---- 64

```

```

QY 158 GGTATATATGACCTCGAGGGCGGTGTGAACACAGAAAGCGCATGAACCGCGATGGCGG 217
Db ----- 64
QY 218 ATGATCCCGAGGGC---GGCGGTGGAACACAGAGCT-----CGCAGCAGCCG 265
Db -----ProValylLeuylGlyylGluAsnSerProGlyLysnAspPro 82
QY 266 CAGCCCCCGGAGGAGCGCCGAGCGCCATGAGGAGGTCCGACCGCCGAGAACATG 325
Db SerLysProGlyLysLysAspAspLeuSerProAsnSerLysAspGlyLysLeu 102
QY 326 CAGCCACGAATCTCGCGCAGAACTTACGCTGTTCAGAGTGAGGAGCTGGAAGTGT 385
Db LysGlnAlaArgSerArgThrAsnPheThrLeuGlnGlnLeuGlnGlnArgLeu 122
QY 386 TTCCGACACACTGAATACCTGATGTGCCACAAAGAGAACTTCCGAAACTAGGT 445
Db PheAspGlnThrHisTyrProAspAlaPheMetArgGlnGlnLeuSerGlnArgLeu 142
QY 446 GTGACTGAAGACAAAGTCGGGTTTGTGTTAAGATAAAGGCGCATGTAGCGACAT 505
Db 143 LeuSerGlnAlaArgValGlnValTyrPheGlnAsnArgAlaGlyLysCysArgLysGln 162
QY 506 CAGAGCAATTA-----ATGCTCCCAATCAATCACTAGCTGTAC 544
Db 163 GluAsnGlnLeuGlnLysGlyLysAspMetLeuCysSerAlaLeuArgLysLeu 180

RESULT 5
Q8AYJ4 PRELIMINARY; PRT; 317 AA.
AC Q8AYJ4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Homeobox transcription factor RAX1.
GN RAX1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxId=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22291598; PubMed=12403708;
RA Chen C.M., Cepko C.L.;
RT "The chicken Rax1 gene plays a role in the initiation of photoreceptor
RT differentiation."
RL Development 129:5363-5375(2002).
DR EMBL; AF420600; AAN32718.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0007275; P:development; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR003654; Homeo_OAR.
DR InterPro; IPR007104; Paired_homeo.
DR Pfam; PF00046; homeobox; 1.
DR Pfam; PF03826; OAR; 1.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1..1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
DR PROSITE; PS50803; OAR; 1.
SQ SEQUENCE 317 AA; 34022 MW; E279B75ECD45FC9D CRC64;

```

```

US-09-867-753-1 (1-577) x Q8AYJ4 (1-317)
QY 119 GCAGAGGCCATGTTGGCCAGAGAGCTCCAGGCTCATG-----GGTATATGAA 169
Db 11 AlaGlnGlyAlaPheSerLeuSerAlaProAlaAlaArgSerProGlyLysnProSer 30
QY 170 CCTGAGGGCGGTGTGAACACAGAGAACCGCATGAACCGCATGCGGCATGATC----- 223
Db ArgLeuHisSerIleGlnAlaIleLeuGlyPheThrLysAspAlaLeuLeuGlyPro 50
QY 224 -----CCGAGGGCGCGGTGGAACACAGAG----- 250
Db 51 PheGlnProAspGlyAlaGlySerAlaLysGlnAlaAlaAspLysArgGlyProArg 70
QY 251 -----CTCGAGCAGACCGCGCCCGCCGCGAGAG----- 283
Db 71 HisCysLeuProLysGlyProAlaGluProProAlaGlnHisGlnGlnLysArgPheGln 90
QY 284 -----CCGCGCCAGCGCGCATGAGGCT 307
Db 91 GluProTyrCysProGlySerAlaSerProGlnLeuProAlaGlyAspGlyLysArgLys 110
QY 308 CCGCAGCCCGAACAATGAGCCA-----CGAATCGCGCAGCAAGTTCAAG 355
Db 111 LysProSerAspGlnGlnGlnProLysLysHisAlaArgAsnArgThrPheThr 130
QY 356 CTGTTGACAGTACGAGAGCTGGAAAGTGTTCGACACACTCAATACCTGATGTCGCC 415
Db 131 ThrTyrGlnLeuHisGlnLeuGlnArgAlaPheGlnLysSerHisTyrProAspValTyr 150
QY 416 ACAGAGAGGAATCTGCCGAAATCTTAGGTGTGACTGAACAGAAAGTCGGGTTGTTT 475
Db 151 SerArgGlnGlnLeuAlaMetLysValaLeuLeuProGlnValaArgValGlnValTyrPhe 170
QY 476 AAGAATAAAGGCGCAGATGTAGGCGCATCAGAGAGATTAATGCTGCCAATGAACA 535
Db 171 GlnAsnArgArgAlaLysTyrPheArgGlnGlnLysLeuGlnValSerSerMetLysLeu 190
QY 536 CCGTGTGACCCA 547
Db 191 GlnAspSerPro 194

RESULT 6
Q8T8C1 PRELIMINARY; PRT; 391 AA.
AC Q8T8C1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Aristaless protein.
GN GBAL.
OS Gryllus bimaculatus (Two-spotted cricket).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Orthopteroidea; Orthoptera; Ensifera; Gryllidae; Gryllinae;
OC Gryllus.
OC NCBI_TaxId=6999;
RN [1]
RP SEQUENCE FROM N.A.
RA Miyawaki K., Inoue Y., Mito T., Matsushima K., Shimmyo Y., Ohuchi H.,
RA Noji S.;
RT "Expression patterns of aristaless in appendage morphogenesis of the
RT hemimetabola, Gryllus bimaculatus (the cricket).";
RL Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; AB071147; BAB85815.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0007275; P:development; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR003654; Homeo_OAR.
DR InterPro; IPR00047; HTH_lamdrepressr.

```

[illegible]

RC TISSUE=Embryo;
RA Sone K., Takeshima K., Takahashi T., Takabatake T.;
RL Dev. Genes Evol. 207:147-155(1997).
RM [2]
RN
RP SEQUENCE OF 129-175 FROM N.A.
RX MEDLINE=56136334; PubMed=8573168;
RA Takabatake T., Takahashi T.C., Inoue K., Ogawa M., Takeshima K.;
RT "Activation of two Cypnos genes, fork head and sonic hedgehog, in
RL animal cap explants,"
CC Biochem. Biophys. Res. Commun. 218:395-401(1996).
-1- SIBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; AB001938; SAA24005.1; -.
DR EMBL; D64022; BAA10897.1; -.
DR HSSP; P06601; 1FTJ.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; P:transcription factor activity; IEA.
DR GO; GO:0006555; P:regulation of transcription, DNA-dependent; IEA
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR007104; Paired_homeo.
DR Pfam; PF00046; homeobox; 1.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SMO0389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
KW SEQUENCE 248 AA; 27667 MW; 96BDA7D61DB59D6 CRC64;

Alignment Scores:	
Pred. No.:	1.05e-09
Score:	200.00
Percent Similarity:	41.24%
Best Local Similarity:	28.87%
Query Match:	19.05%
DB:	13
Length:	248
Matches:	56
Conservative:	24
Mismatches:	52
Indels:	62
Gaps:	6

US-09-867-753-1 (1-577) X 012952 (1-248)

```

QY      2  CGAACATCAGGCGCCTCCAGCAATGGCGCGTTGGCTGCGT---CAGACACCGCTTTCTAC  58
      |||||:::|||||
DB      58  ProThrlaetIleuProThrlleuIaetIleSerIleValIglIYraSgIleIYrTr  77
QY      59  TGCGTGAAGTATACAGAGGTAAAAATTAAGCCCCCAACCTCAGCTGGGGGAGCATCAAGC  118
      |||:::|||||
DB      78  -----TyrIglIleuIhIleuGlInPro-----  85
QY      119  GCAGAAAGGCCATTTGGCCAGAGAGCTCCAGGCGCTCATGGTAAATATGAACTTGAGAGGC  178
      |||||:::|||||
DB      86  -----ProglYlAlaProglYcyCySglYlAlaMetGlInProleuGlY  99
QY      179  GGTGTGAACCAAGAAACGGCATGAACCGCATGGCGGCGATGATCCCGAGAGCGGCGGT  238
      :::
DB      100  Ala-----GInGIncySerYcYValProAlaProSerSer  111
QY      239  GGAACCAAGAGCCTTGCGGACGACGCCGCGAGCCCCCGGAGAGAGCCGAGCCCAAGCGGC  296
      |||:::|||||
DB      112  GlY-----TyrAspGlYSerSerVal  119
QY      299  ATGAGAGGTCGCGAGCGCCGAGAAACATGCAGCCA  331
      :::
DB      120  LeuMetSerProMetProHISglInMetMetProTYrMetAenValIglIYThleuSerArg  133
QY      332  -----CGAACTCCGCGGACGAG  345
DB      140  SerGlInleuGlInleuIeuaenGlInleuIhIScySArgArgIYsArgArgIhISArgThIle  155
QY      350  TTCACGCTGTTCAGAGTGAAGAGAGCTCGAAAGTGTTTTCCACACACTCAATATACCCTGAT  409
      |||||:::|||||
DB      160  PheThrAspIleuGlInleuGlInIaIeIeugIhISleuPheGInGInIuThIYrProAsp  179
QY      410  GTGCCCAAGAAAGGGAACCTGCCGAAAACCTTAAGTGTGACTGAATGAAGCAAAAGTCCGGGTT  465
      |||||:::|||||
DB      180  ValIglIYhTrIrgIuGlInleuIaIaArgValIhISleuAdrgGInGInIuIYValGluVal  199

```

QY 470 TGGTTTAAGATTAATAAGGCCAGATGTAGCCAGCATCAGAGA 511
 DB 200 TTPhelyAAsnArgAlaYstTTPArgGlnLysArg 213
 RESULT 8
 ID 088933 PRELIMINARY; PRT; 382 AA.
 AC 088933;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Homeobox protein SPXL.
 GN ESX1 OR EPX.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97398449; PubMed=9256347;
 RA Bratford W.W., Zhao G.Q., Valerius M.T., Weinstein M.,
 RA Birkenmeier E.H., Rowe L.B., Potter S.S.;
 RT "Spxl, a novel X-linked homeobox gene expressed during
 RT spermatogenesis.";
 RT Mech. Dev. 65:87-98 (1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97392770; PubMed=9245514;
 RA Li Y., Lemaire P., Behringer R.R.;
 RT "Esxl, a novel X chromosome-linked homeobox gene expressed in mouse
 RT embryonic tissues and male germ cells.";
 RT Dev. Biol. 188:85-95 (1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Bratford W.W., Potter S.S.;
 RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SV;
 RA Yan Y.-T., Yang L., Sciavolino P.J., Wang H., Chan D.C.,
 RA Abate-Shen C., Shen M.M.;
 RT "EpX: a novel paired-like homeobox gene expressed in the chorion and
 RT placenta.";
 RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
 CC -I- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 DR EMBL; AF085715; AAC35366.1; -;
 DR EMBL; AF017735; AAD01622.1; -;
 DR HSSP; P06601; IFUL.
 DR TRANSFAC; T03478; -;
 DR MGD; MGI:1096388; Esxl.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001356; Homeobox.
 DR InterPro; IPR000047; HTH_LambdaRepressr.
 DR Pfam; PF00046; homeobox; 1.
 DR PRINTS; PR00031; HTHREPRESSR.
 DR PRODOM; PD000010; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX_1; 1.
 DR PROSITE; PS50071; HOMEBOX_2; 1.
 KW DNA-binding; Homeobox; Nuclear protein.
 SQ SEQUENCE 382 AA; 43540 MW; EC92301A84DA6175 CRC64;
 Alignment Scores:
 Pred. No.: 2.05e-09 Length: 382
 Score: 197.00 Matches: 59
 Percent Similarity: 44.51% Conservative: 18
 Best Local Similarity: 34.10% Mismatches: 72
 Query Match: 18.76% Indels: 24
 DB: 11 Gaps: 5

US-09-867-753-1 (1-577) x 088933 (1-382)
 QY 71 TACCAGTAAATAATAGCCACACCTGAGTGGGGAGCATCAGCGCAGAGGCCAT 130
 DB 84 TGTGInGInProGInGInGInPheGInProSerArgGInGInAlaAlaPro----- 100
 QY 131 GTTGGCCAAAGAGCTCCAGGCTCATGGTAAATATGAACCTGAGGGCGGTGAACAC 190
 DB 101 VALAAGInAlaPProGInAlaTTPAsnGInAsnGInGInGInGInGInGIn 120
 QY 191 GAGAACGGCATGACCGGATGGCGG----- 217
 DB 121 SerAsnAlaGInLeuGInGInGInAlaAspAlaAlaProValArgGInSerLeuMetArgPro 140
 QY 218 --ATGATCCCGAGGGCGCGCGTGAACCCAG-----GAGCCTCGGACGACG 262
 DB 141 LeuMetGInProValAlaGInSerSerProGInProLeuProAlaAsnProLeuGInAla 160
 QY 263 CCGCAGACCCCG-----CCGAGAGAGCCGCGCCAGCGGACATGAGGTCG 310
 DB 161 ProGInGInProGInGInGInGInGInGInGInGInGInGInGInGInGInPro 180
 QY 311 CAGCCCGAGACATGACGACGACACTCGCGCAGAGTTCACGCTGTGGAGGAGGAG 370
 DB 181 GInGInGIn--ProLysProArgArgTyrArgTyrCysPheTrnProLysGInGIn 199
 QY 371 GAGCTGAAAGTGTTCCTCCGACACACTCAATACCTGATGTGCCACAGAGGAAGT 430
 DB 200 GInLeuGInAlaPhePheGInArgValGInTyrProAspLeuPheAlaArgValGInLeu 219
 QY 431 GCCGAAACTTGGTGTGACTGAAGACAAAGTCCGGTGTGTTAAGATTAAGAGGCGC 490
 DB 220 AlaArgArgLeuGInGInLeuProGInProArgValGInValTTPheGInAsnArgArgAla 239
 QY 491 AGATGTAGGCGACATCAGAGAGATTAATGCTCGGCAAT 529
 DB 240 LysTTPArgArgLeuArgArgAlaGInAlaPheArgAsn 252
 RESULT 9
 ID 092203 PRELIMINARY; PRT; 387 AA.
 AC 092203;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Homeodomain protein EPX (Fragment).
 GN ESX1 OR EPX.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SV;
 RX MEDLINE=20079279; PubMed=10611245;
 RA Yan Y.-T., Stein S.M., Ding J., Shen M.M., Abate-Shen C.;
 RT "A Novel PE/PN Motif Inhibits Nuclear Localization and DNA Binding
 RT Activity of the ESX1 Homeoprotein.";
 RL Mol. Cell. Biol. 20:661-671 (2000).
 CC -I- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 DR EMBL; AF017734; AAD01621.1; -;
 DR HSSP; P06601; IFUL.
 DR MGD; MGI:1096388; Esxl.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001356; Homeobox.
 DR InterPro; IPR000047; HTH_LambdaRepressr.
 DR InterPro; IPR007104; Paired_homeo.
 DR Pfam; PF00046; homeobox; 1.
 DR PRINTS; PR00031; HTHREPRESSR.
 DR PRODOM; PD000010; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.

DR PROSITE; PS00027; HOMEBOX 1; 1.
 DR PROSITE; PS0071; HOMEBOX 2; 1.
 KW DNA-binding; Homeobox; Nuclear protein.
 FT NON TER 1
 SQ SEQUENCE 387 AA; 44170 MW; 0D6A6E4C36CF9CD3 CRC64;

Alignment Scores:
 Pred. No.: 2, 06e-09 Length: 387
 Score: 197.00 Matches: 59
 Percent Similarity: 44.51% Conservative: 18
 Best Local Similarity: 34.10% Mismatches: 72
 Query Match: 18.76% Indels: 24
 DB: 11 Gaps: 5

US-09-867-753-1 (1-577) x Q9Z2U3 (1-387)

OY 71 TACGAGTAAATAAGCCCACTCAGTGGGGGAGCATCAAGCCGAGAGGCCAT 130
 |||||
 DB 89 TyrgIngluProgluGlyPhegluProserArglyGluAlaAlaPro----- 105

OY 131 GTTGCCCAAGAGCTCCAGGCTCATGGTATATGAAACCTGAGGGCGGTGTAACAC 190
 |||||
 DB 106 ValAlaGluAlaProGlnAlaTTrpAnnglyAnngluAnleuGlyGlyPheleuGlu 125

OY 191 GAGAACGGCATGACCGCGATGGCGC----- 217
 |||||
 DB 126 SerAenAlaGlnleuGlyGluAlaAspAlaAlaProValArgGlnSerleuMetArgPro 145

OY 218 --ATGATCCCGAGGGCGCGGTGMAACAG-----GAGCTCGGCGAGCAG 262
 |||||
 DB 146 LeuMetGlnProValAlaGlnSerSerProGlnProleuProAlaAsnProleuGlnAla 165

OY 263 CCGGAGCCCGC-----CCGAGAGCGCGCCAGCGGCCCATGAGGGTCCG 310
 |||||
 DB 166 ProGlnGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 185

OY 311 CAGCCCGAAGACATGACGACCAACTCGCGCAGCAAGTTGCGTGGAGGGTGGAG 370
 |||||
 DB 186 GlnGlnGln---ProlyeProArgArglyTyArgIleCySPheThProIleGlnleuGln 204

OY 371 GAGCTGGAAGGTGTTTCCGACACACTCAATACCTGATGTGCCCAAGAAAGGAATT 430
 |||||
 DB 205 GluLeuGlnAlaPhePheGlnArgValGlnTyTrProAspLeuPheAlaArgValGluLeu 224

OY 431 GCCGAAACTTGGTGTGACTGAAAGCAAAAGTGGGTTGTTTGAATTAAGGGCC 490
 |||||
 DB 225 AlaArgArgleuGlyLeuProGlnProArgValGlnValTTrPheGlnAsnArgArgAla 244

OY 491 AGATGTAGCGCATCAGAGAGATTAATGCTCGCCAAAT 529
 |||||
 DB 245 LysTrpArgArgleuArgArgAlaGlnAlaPheArgAsn 257

RESULT 10
 O54817 PRELIMINARY; PRT; 314 AA.
 AC O54817
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Paired-like homeodomain containing protein.
 GN ESX1 OR ETV.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Swiss;
 RX MEDLINE=97392770; PubMed=9245514;
 RA Li Y., Lemaire P., Behringer R.R.;
 RT "Esx1, a novel X chromosome-linked homeobox gene expressed in mouse
 extraembryonic tissues and male germ cells.";
 RL Dev. Biol. 188:85-95(1997).

CC -I- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 DR EMBL; AF004211; AAB94670.1; -.
 DR HSSP; P06601; 1FJL.
 DR TRANSFAC; T03474; -.
 DR MGD; MGI:1096388; Esx1.
 DR GO; GO:0005634; C:nucleus, IEA.
 DR GO; GO:0003700; F:transcription factor activity, IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent, IEA.
 DR InterPro; IPR01356; Homeobox.
 DR InterPro; IPR000047; HTH_lambdarepressr.
 DR InterPro; IPR007104; Paired_homeo.
 DR Pfam; PF00046; homeobox; 1.
 DR PRINTS; PR00031; HTHREPRESSR.
 DR ProDom; PD000010; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX_1; 1.
 DR PROSITE; PS0071; HOMEBOX_2; 1.
 KW DNA-binding; Homeobox; Nuclear protein.
 SQ SEQUENCE 314 AA; 35969 MW; 72568F81DA4AA246 CRC64;

Alignment Scores:
 Pred. No.: 2, 49e-09 Length: 314
 Score: 196.00 Matches: 59
 Percent Similarity: 44.51% Conservative: 18
 Best Local Similarity: 34.10% Mismatches: 72
 Query Match: 18.67% Indels: 24
 DB: 11 Gaps: 5

US-09-867-753-1 (1-577) x O54817 (1-314)

OY 71 TACGAGTAAATAAGCCCACTCAGTGGGGGAGCATCAAGCCGAGAGGCCAT 130
 |||||
 DB 16 TyrgIngluProgluGlyPhegluProserArglyGluAlaAlaPro----- 32

OY 131 GTTGCCCAAGAGCTCCAGGCTCATGGTATATGAAACCTGAGGGCGGTGTAACAC 190
 |||||
 DB 33 ValAlaGluAlaProGlnAlaTTrpAnnglyAnngluAnleuGlyGlyPheleuGlu 52

OY 191 GAGAACGGCATGACCGCGATGGCGC----- 217
 |||||
 DB 53 AlaAenAlaGlnleuGlyGluAlaAspAlaAlaProValArgGlnSerleuMetArgPro 72

OY 218 --ATGATCCCGAGGGCGCGGTGMAACAG-----GAGCTCGGCGAGCAG 262
 |||||
 DB 173 LeuMetGlnProValAlaGlnSerSerProGlnProleuProAlaAsnProleuGlnAla 92

OY 263 CCGGAGCCCGC-----CCGAGAGCGCGCCAGCGGCCCATGAGGGTCCG 310
 |||||
 DB 93 ProGlnGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 112

OY 311 CAGCCCGAAGACATGACGACCAACTCGCGCAGCAAGTTGCGTGGAGGGTGGAG 370
 |||||
 DB 113 GlnGlnGln---ProlyeProArgArglyTyArgIleCySPheThProIleGlnleuGln 131

OY 371 GAGCTGGAAGGTGTTTCCGACACACTCAATACCTGATGTGCCCAAGAAAGGAATT 430
 |||||
 DB 132 GluLeuGlnAlaPhePheGlnArgValGlnTyTrProAspLeuPheAlaArgValGluLeu 151

OY 431 GCCGAAACTTGGTGTGACTGAAAGCAAAAGTGGGTTGTTTGAATTAAGGGCC 490
 |||||
 DB 152 AlaArgArgleuGlyLeuProGlnProArgValGlnValTTrPheGlnAsnArgArgAla 171

OY 491 AGATGTAGCGCATCAGAGAGATTAATGCTCGCCAAAT 529
 |||||
 DB 172 LysTrpArgArgleuArgArgAlaGlnAlaPheArgAsn 184

RESULT 11
 O54817 PRELIMINARY; PRT; 408 AA.
 AC O54817
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)

AL gene product (RE68460p).
 GN AL OR CG3935.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydriidae; Drosophilidae; Drosophila.
 NCBI_Taxid=7227;
 (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkely;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amaratilake P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolhakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Broksrein P., Brotler P.,
 RA Burks C.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostel C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeigbam C.,
 RA Jitali M., Kalish F., Kapen G.C., Kravitz S., Kulp D., Lai Z.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson A.,
 RA Merkulov G., Mlshina N.V., Mobarry C., Morris J., Mostrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazolo J., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Svirer E., Spralling A.C., Stapleton M., Strong R., Sun B.,
 RA Svirer E., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 (12)
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkely;
 RA Stapleton M., Broksrein P., Hong L., Agbayani A., Carlson J.,
 RA Change M., Chavez C., Dorsett V., Dresnek D., Fafan D., Frise E.,
 RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
 RA Miranda A., Mungall C.J., Nunoo J., Pacle J., Paragas V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
 RA Celinker S.;
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBS databases.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 DR EMBL, AEO03589, AAF51505.1; -.
 DR EMBL, AY121696, AAM52023.1; -.
 DR PIR: A40685, A40685.
 DR HSSP: P06601, 1FUT.
 DR FlyBase; FBgn0000061, al.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0007275; P:development; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001356, Homeobox.
 DR InterPro; IPR003654, Homeo_OAR.
 DR InterPro; IPR000047; HTH_LambdaRepressr.
 DR InterPro; IPR007104; Paired_homeo.

DR Pfam; PF00046, homeobox, 1.
 DR Pfam; PF03826, OAR, 1.
 DR PRINTS; PR00031; HTHREPRESSR.
 DR Prodom; PD000010, Homeobox, 1.
 DR PROSITE; PS00027, HOMEBOX_1, 1.
 DR PROSITE; PS0071, HOMEBOX_2, 1.
 DR PROSITE; PS0803, OAR, 1.
 KW DNA-binding; Homeobox; Nuclear protein.
 SQ SEQUENCE 408 AA; 43657 MW; 7641316AA77BD8D CRC64;
 Alignment Scores:
 Pred. No.: 2,55e-09 Length: 408
 Score: 196.00 Matches: 47
 Percent Similarity: 55.17% Conservative: 17
 Best Local Similarity: 40.52% Mismatches: 25
 Query Match: 18.67% Indels: 27
 DB: 5 Gaps: 3
 US-09-867-753-1 (1-577) x Q9VPP1 (1-408)
 QY 244 CGAGGAGCCTCGGACAGCCGAGCCGCC-----GGA 279
 Db 31 ProlglyserSerAlaAlaSerAlaGlyAlaAlaLeuThValSerMetSerValSerGly 50
 QY 280 GGAGCCGCGCCAGCGCGCCATGAGAGGTCGCA----- 312
 Db 51 GlyAlaProSerGlyAlaSerGlyAlaSerGlyGlyThrAsnSerProValSerAspGly 70
 QY 313 -----GCCGGAACATGACCCACCAACTCGG-----CGCAC 345
 Db 71 AsnSerAspCysGluAlaAspGluTyrAla-ProlysArgLysGlnArgArgTyrArgTh 90
 QY 346 GAAGTCAGCGCTGTCGAGGAGGCTGGAAAGTGTTCCTCCGACACATCAATACCC 405
 Db 90 ThrPheThrSerPheGlnIleuGluIleuGluLysAlaPheSerArgThHisTyrTr 110
 QY 406 TGAATGCCACACAGAGGAATCTGCCGAAACTTAGGTGACTGAAGACAAAGTCGC 465
 Db 110 AspValPheThrArgGluGluLeuLysIleGlyLeuThGluAlaArgIleG 130
 QY 466 GGTTCGTTTAAGATATAAAGGCGGAGATGTGGCAGACTCAGAGA 511
 Db 130 nValTrpPheGlnAsnArgArgAlaLysTrpArgLysGlnGluLys 145
 RESULT 12
 Q9V7M5 PRELIMINARY; PRT; 328 AA.
 AC Q9V7M5;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Orthopedia protein.
 GN OTP OR OTP.
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 NCBI_Taxid=7955;
 (1)
 RP SEQUENCE FROM N.A.
 RA Del Giacco L., Di Benedetto B., Duga S., Diani S., Cotelletti F.,
 RT "Isolation of the mRNA encoding Otp (Orthopedia) in the zebrafish,
 RT Danio rerio.";
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBS databases.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 DR EMBL, AF071496, AAD42021.1; -.
 DR HSSP: P06601, 1FUT.
 DR ZFIN; ZDB-GENE-990708-7; otp.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0007275; P:development; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001356, Homeobox.

DR InterPro; IPR003654; Homeo_OAR.
 DR InterPro; IPR000047; HTH_lambdarepressor.
 DR InterPro; IPR007104; Paired_homeo.
 DR Pfam; PF00046; homeobox_1.
 DR Pfam; PF03826; OAR; 1.
 DR PRINTS; PR00024; HOMEBOX.
 DR PRINTS; PR00031; HTHREPRESSR.
 DR ProDom; PD000010; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX_1; 1.
 DR PROSITE; PS0071; HOMEBOX_2; 1.
 DR PROSITE; PS00803; OAR; 1.
 DR DNA-binding; Homeobox; Nuclear protein.
 KW SEQUENCE 328 AA; 35488 MW; 7CCABCA89011C3EC CRC64;

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	DB:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
5	79e-09	192.00	42.05%	13	13	328	60	22	65	48	8

US-09-867-753-1 (1-577) x Q9W7M5 (1-328)

QY 2 CCAACATCAGG-----GCTCCAGCATGGCGCTTCGCTCCAC 43
 DB 18 ProhomerGlyLeuGlnLeuH1sTTPMetLysAspAlaAlaLeuLeuValHis 37
 QY 44 GACACCGTGTCTACTCCTGAGTATACAGTAAATAAGCCCACTCAGCTG 103
 DB 38 ArgGluAlaLeuLecys-----ArgLeu 45
 QY 104 GGGGAGCATCAAGGCGAAGGCGCTGTGCG----- 136
 DB 46 GlyGlyThrAspSer-----GlyHisProGlyAspLeuThrSerAlaThrGluThrVal 63
 QY 137 CAAGAGCTCCAGGCTCATGGGTATATGAACCTGAGGCGGCTGTGAACCAAGAG 196
 DB 64 GluGlyThrThrLeuLeuProGlyGluGluLeuSerAsnGlyLysSerHn--ProHn 82
 QY 197 GGCATGAACCGCGATGGCGCATGATCCCGAGGCGCGGTGGAACCAAGAGCCTCG 256
 DB 83 GlyMetGln-----GlnValAsnAlaLysAspGlnGluLysGln 95
 QY 257 CAGCAGCCGAGCCCGCCGAGAGAGCCGCGCCAGCGCCATGAGAGGCTCCGAGCC 316
 DB 96 GlnGln-----GlnAsnSerAsnGlnThrGlyGlyGlnGln 107
 QY 317 GAGAACATGACGACGAACTGGCGGACGAAAGTTGACGCTGTGGAGGAGAGAGCTG 376
 DB 108 AsnGlnGlnLysGlnLysArgHisArgHisArgHisProAlaGlnLeuGlnLeu 127
 QY 377 GAAAGTGTTCGACACACTCATACCTGATGTGCCCAAGAGAGGAACTTGGCGAA 436
 DB 128 GluArgSerPheAlaLysThrHisTyrProAspIlePheMetArgGluGluLeuAlaLeu 147
 QY 437 AACTTGGTGTGACTGAAGACAAAGTCGGGCTTGTGTTAAGATTAAGGCGCAGATGT 496
 DB 148 ArgIleGlyLeuThrLeuSerArgValGlnValTTPPheGlnAsnArgArgAlaLysTTP 167
 QY 497 AGGCGCATCGAGAGAAATTAATGCTCCGCAATGAAGTACGTCT 541
 DB 168 LysLysArgLysLys-----ThrThrAsnValPheArgAla 179

RESULT 13

Q8BPD6 PRELIMINARY; PRT; 227 AA.
 AC Q8BPD6; 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
 DE Placenta specific homeobox 1.

GN PSX1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OK NCBI_Taxid=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Placenta, and Extraembryonic tissue;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The PANTOM Consortium,
 RT the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573 (2002).
 DR EMBL; AK076186; BAC36240.1; -.
 DR MGD; MGI:1202888; Psx1.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; P:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001356; Homeobox.
 DR Pfam; PF00046; homeobox; 1.
 DR ProDom; PD000010; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00071; HOMEBOX_2; 1.
 DR SEQUENCE 227 AA; 25544 MW; 0FEFF42ADB5F68F04 CRC64;

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	DB:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
8	53e-09	190.00	46.39%	11	11	227	55	22	57	32	5

US-09-867-753-1 (1-577) x Q8BPD6 (1-227)

QY 104 GGGGAGCATCAAGGCGAAGGCGCATGTGGCCAGAGACTCCAGGCTCATGGTAAAT 163
 DB 54 GlyGluPheAspGlnGlyGluGluValAlaGlyGlyGluGlnAlaGln 73
 QY 164 ATGAACCT-----GAGGCGGTGTAACCAAGAGCGGATG 202
 DB 74 GluGluProAlaProLeuSerProAlaGlnGluAlaThrGlyGlyGluGluGlu 93
 QY 203 AACCGGATGGGGCATG----- 220
 DB 94 AsnLysGlnGlyGluMetGlnGlyArgHisAlaGlyAspGlyAlaSerSerGluAsp 113
 QY 221 -----ATCCCGAGAGGCGCGGTGGAAC-----CAGAGCTTCGCGAGCGCGAG 268
 DB 114 AspSerIleLeuGlnGluGlyGlyGluLeuIleAspGlnGlnProProGlnGlnAla 133
 QY 269 CCCCCCGGAGAGCCGCGCCAGCGCCATGAGAGGTCGCAAGCCGAGAAC----- 322
 DB 134 AlaSerProAsp-----SerIleArgAsnProHisValLeuAsnArgLeu 148
 QY 323 ATGACGCAAGAACTCGGCGGACGAAGTTCAAGCTGTTCGAGAGGAGGAGGAAAT 382
 DB 149 AlaGlnLeuArgTyrArgArgThrArgPheThrHisPheGlnLeuHisAspLeuGluArg 168
 QY 383 GTTTCGACACATCAATACCTGATGCGCCACAGAGAGGAACTTGGCGAAACTTA 442
 DB 169 LeuPheGlnGlnThrArgTyrProSerLeuAlaArgArgAspLeuAlaArgTyrMet 188
 QY 443 GGTGTGACTGAAGACAAAGTCGGGCTTGTGTTAAGATTAAGGCGCAGATGAGCGCA 502
 DB 189 GlyValAspGluCysAspValGlnAsnTTPPheArgMetArgArgAlaLeuPheGlnArg 208
 QY 503 CATCAGAGGAATTAATG 520
 DB 209 AsnArgArgValLeuMet 214

RESULT 14

RA Dodaio K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.
 RA Ferreira S., Frisè E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hopkins R.A., Hostin D., Howland T.J.,
 RA Ibgweg C., Jalili M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA Mcintosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Patel J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pltman G.S., Puri V., Richards S., Scheeler F.,
 RA Stepieton M., Strong R., Svirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri U.S., Smith H.O., Venter J.C., Rubin G.M.,
 RT "Sequencing of Drosophila melanogaster genome."
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Miura S., Crosby M.A., Matthews B.B., Bayraktaroglu I., Campbell K
 RA Hirdcey P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Bertman B., Carlson J.W., Celisner S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,
 RT "Annotation of Drosophila melanogaster genome."
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celisner S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RL [6]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -1- SIMILARITY: CONTAINS 1 PAIRED BOX DOMAIN.
 DR EMBL; AY069177; AAL39322.1; -
 DR EMBL; AE006354; AAF49919.2; -
 RA FlyBase; FBGN0036285; toe.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; P:transcription factor activity; IEA.
 DR GO; GO:0007275; P:development; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR GO; GO:0006350; P:transcription; IEA.
 DR InterPro; IPR001356; Homeobox.
 DR InterPro; IPR001523; Paired box.
 DR InterPro; IPR007104; Paired_homeo.
 DR Pfam; PF00046; homeobox; 1.
 DR Pfam; PF00292; PAX; 1.
 DR PRINTS; PR00027; PAIRBBOX.
 DR ProDom; PD000010; Homeobox; 1.
 DR SMART; SMO0389; HOX; 1.
 DR SMART; SMO0351; PAX; 1.
 DR PROSITE; PS00027; HOMEBOX_1; 1.
 DR PROSITE; PS00071; HOMEBOX_2; 1.
 KW DNA-binding; Developmental protein; Homeobox; Nuclear protein;
 KW Paired box; Transcription; Transcription regulation.
 CO SEQUENCE 640 AA; 66525 MW; 306232EAF4BBB32B CRC64;

Alignment Scores:	
Pred. No.:	9.34e-09
Score:	190.00
Percent Similarity:	43.86%
Best Local Similarity:	32.16%
Query Match:	18.10%
DB:	5
Length:	6406
Matches:	55
Conservative:	20
Mismatches:	60
Indels:	36
Gaps:	7

US-09-867-753-1 (1-577) x Q8TOM4 (1-640)

Qy	92	ACACCTTCAGCTGGGGGACATCAGCAGGAGGCGATCTTCAGGC	151
		:::	
Db	279	ThrProProGlyGlyGlyProAlaGlyAlaGlyGlyAlaLeuInProGlyGlySerGly	298
		:::	
Qy	152	CTCAGG-----GCTAATATG-----AACCTT-----GAGGCGCGATGTG	184
Db	299	SerSerTyrGlySerAspIlyAsnMetSerSerAsnIProAsnSerSerAsnSerAsnThr	318

QY	188	AACCCAGAGAAACGGCATGAACCGCATGGCGGC-----	21.7
		:::	
Db	319	ThrlsSerAmGlyHhlsnThrnsmberldYserGlyCyGslYabpSerSerAlaGly	33.8
QY	218	-----ATGATCCCGCAGGCGCGCGGTGGAAACAGAGACCT	25.3
		:::	
Db	339	SerGlyYrGleuSerLeuProAlaLeuSerProAlaSerSerGlySer-----	35.3
QY	254	CGGCAGCAGCCGCGAGCCCGCCGCGAGAGCCGCGCCAGCGGCATGAGAGGTCGCGCAG	31.3
		:::	
Db	354	ArgAspSerArgSerProAlaAlaSerAlaAlaSerMetIleAspIleGluGlyGluAsp	37.3
QY	314	CCGAGAAACATG-----CAGCCAGCAACTCGG-----CGCAGAAAGTTACAGCTG	35.8
Db	374	SerGlnSerGlnAspSerSerAspGlnProGlySphaArgAlaGlnAlaGlnThrPheSerPro	39.3
QY	359	TTGCAGGTGAGAGAGCTGGAAAGTTTTCGCAACAACACTCAATACCTGTATGTCGCCACA	41.8
		:::	
Db	394	GluGlnLeuAspGluLeuGlnIuyluSerGluPheAspIlySerHisTyrProCysValAsnThr	41.3
QY	419	ACAGAGGAACCTGCCGAAACCTTAGCGTGACTGAAGACAAAGTCGGGTTTGCTTTAAG	47.8
		:::	
Db	414	ArgGluIuyluSerAlaAlaArgThrAlaLeuSerGluAlaArgValGlnValITPheSer	43.3
QY	479	AATTAAGGCCAGATGTATGCGCATCATCAGAGA	51.1
Db	434	AsnArgArgAlaIuylTrpArgArgHisGlnIuylArg	44.4

Search completed: June 16, 2004, 19:02:26
Job time : 52 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 16, 2004, 18:56:45 ; Search time 20.5 Seconds
(without alignments)
5414.879 Million cell updates/sec

Title: US-09-867-753-1

Perfect score: 1050
Sequence: 1 tccacatcagcgcctccag.....tctacatcgctgtagctagc 577

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+zip.model -DB=xlp
-Q=/cgn2.1/USPTO_spool_p/US09867753/runat_16062004_163636_27806/app_query.fasta_1.775
-DB=PIR_78 -QFMT=fastan -SUFFIX=prp -MIMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NCM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09867753 -GCN 1.1 38.0 runat_16062004_163636_27806 -NCP=6 -ICPU=3
-NO MMAP -LARGEBUFFER -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_78:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	196	18.7	384	2	A46403
2	189.5	18.0	399	2	A46422
3	189	18.0	252	1	A54677
4	188	17.9	245	1	A47539
5	187	17.8	256	1	A42768
6	184.5	17.6	217	2	I48902
7	184.5	17.6	245	2	S26076
8	181	17.2	419	1	S70617
9	180	17.1	333	1	JW0097
10	179.5	17.1	326	2	A45452
11	179.5	17.1	362	2	T24046
12	178	17.0	185	2	A55882
13	176.5	16.8	290	2	I49265
14	176	16.8	245	2	I51226

15	176	16.8	247	1	JC6540	placenta specific-
16	176	16.8	280	2	I48713	Phox2 homeodomain
17	176	16.8	370	2	UC6130	paired box transcr
18	176	16.8	479	1	S15031	paired box transcr
19	176	16.8	612	2	A54282	reversed polarity
20	173	16.5	80	2	A26332	homeotic protein B
21	173	16.5	346	2	I48185	gene alix protein
22	173	16.5	427	1	A43698	paired box transcr
23	172.5	16.4	613	1	A26062	paired box segment
24	172.5	16.4	798	2	S20881	homeotic protein P
25	172	16.4	449	1	B43698	paired box transcr
26	172	16.4	520	2	S78502	paired box transcr
27	171.5	16.3	243	1	B47668	homeotic protein G
28	171.5	16.3	243	2	I51424	homeotic protein G
29	171.5	16.3	363	2	UC7750	CVC domain-contain
30	168.5	16.0	242	2	A43904	homeotic protein G
31	168.5	16.0	278	2	A56570	homeobox protein D
32	167.5	16.0	283	2	I50112	DLX4 homeodomain P
33	167	15.9	288	2	JS0659	homeotic protein H
34	167	15.9	319	2	T18786	hypothetical prote
35	167	15.9	326	2	A47523	cartilage homeop
36	166.5	15.9	284	2	S60250	mab-18 protein (tr
37	166	15.8	296	2	S60251	paired type homeo
38	166	15.8	314	2	JC5273	homeodomain protei
39	166	15.8	318	2	S52424	homeodomain protei
40	166	15.8	319	2	T43635	homeobox protein H
41	166	15.8	373	2	A47234	homeotic protein s
42	165.5	15.8	288	2	S27842	homeotic protein o
43	165.5	15.8	671	1	A35912	paired box transcr
44	164	15.6	491	2	S60252	transcription fact
45	163.5	15.6	298	2	I51410	

ALIGNMENTS

RESULT 1
A46403
transcription factor with prd-type homeo domain and Pro/Gln-rich domain al - fruit fly (I
C/species: Drosophila melanogaster
C/date: 21-Sep-1993 #sequence_revision 25-Apr-1997 #text_change 15-Oct-1999
C/accession: A46403
R/Schneitz, K.; Spielmann, P.; Noll, M.
Genes Dev. 7, 114-129, 1993
A>Title: Molecular genetics of aristales, a prd-type homeo box gene involved in the mori
A/reference number: A46403; MUID:93138380; PMID:8093690
A/status: preliminary
A/molecule type: nucleic acid
A/residues: 1-384 <SCH>
A>Note: sequence extracted from NCBI backbone (NCBIN:123442, NCBIPI:123444)
C/superfamily: unassigned homeobox proteins; homeobox homology
C/keywords: DNA binding; homeobox; nucleus; transcription regulation
F:86-142/Domain: homeobox homology <HOX>

Alignment Scores:

Pred. No.: 6.19e-09 Length: 384
Score: 196.00 Matches: 47
Percent Similarity: 55.17% Conservative: 17
Best Local Similarity: 40.52% Mismatches: 25
Query Match: 18.67% Indels: 27
DB: 2 Gaps: 3

US-09-867-753-1 (1-577) x A46403 (1-384)
QY 244 CCAGAGCCTCGCAGACGACGCCGCCGCC-----GCA 279
DB 31 ProGlySerSerAlaAlaSerAlaGlyAlaAlaLeuThrValSerMetSerValSerGly 50
QY 280 GAGAGCGCCGCGAGCGCGCCATGAGAGGTCGCCA----- 312
DB 51 GlyAlaProSerGlyAlaSerGlyAlaSerGlyGlyThrAsnSerProValSerAspGly 70
QY 313 -----GCCGAGACATGACGACGACGACACTCGG-----CGCAC 345


```

OY 332 -----CGAAGTCCGCGCAGCAAGTTCACGCTGTTGACAGG 367
DB 148 AengInLeuH1sCyArGrArgLySarGArgH1sArgThrI1lePheThrAspGluInLeu 167
OY 368 GAGGAGCTGGAAAGTGTTCGACACACTCAATACCTGATGTGGCCCAAGAAAGGAA 427
DB 168 G1uAlaLeuG1uAlaLeuPheG1uInG1uThrLyS1rProAspValG1yThrArgGluIn 187
OY 428 CTTGCGCGAAACTTGAAGTGTGACTGAGACCAAGTCCGGTTCGTTTAAGATAAAG 487
DB 188 LeuAlaArgLySValH1sLeuArgG1uInG1uValG1uThrPheLysAsnArgArg 207
OY 488 GCCAGATGTAGCGACATCAGAGA 511
DB 208 AlAluYrTPArGrArgInLySarG 215

RESULT 4
A47539
homeotic protein gooseoid [similarity] - chicken
C/Species: Gallus gallus (chicken)
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 01-Sep-2000
C/Accession: A47539
R: Izpisua-Belmonte, J.C.; De Robertis, E.M.; Storey, K.G.; Stern, C.D.
Cell 74, 645-659, 1993
A/Title: The homeobox gene 'gooseoid' and the origin of organizer cells in the early chick
A/Reference number: A47539; M01D:93364981; PMID:7916659
A/Accession: A47539
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-245 <12P>
A/Cross-references: EMBL:X70471; NID:9402580; PIDN:CAA49897.1; PID:9402581
C/Superfamily: homeotic protein gooseoid; homeobox homology
C/Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:151-207/Domain: homeobox homology <HOX>

Alignment Scores:
Pred. No.: 36-08 Length: 245
Score: 188.00 Matches: 57
Percent Similarity: 40.728 Conservative: 22
Best Local Similarity: 29.388 Mismatches: 51
Query Match: 17.904 Indels: 64
DB: 1 Gaps: 6

US-09-867-753-1 (1-577) x A47539 (1-245)
OY 2 CCAACATCAGGCGCTCCAGCCATGCGCGCTG---CTGCTCCAGCAGCACCGTGTCTAC 58
DB 57 ProG1ySerAlaLeuProAlaValG1yArgSerArgLeuG1yTyRAsnAsnTyRtyr 76
OY 59 TGCCTGAGTGTATACCAAGTAATAAATTAAGCCCACTCAAGTGGGGGACAGATCAAGC 118
DB 77 -----GlyInLeuH1sValAlaThrSer 84
OY 119 GCAGAAAGCCATGTTGGCCAGAGAGCTCCAGGCTCATGGGTAATGAACCTTGAGGCG 178
DB 85 ProValG1y-----ProSerCyS1yG1yAlaValProProLeuG1y 98
OY 179 GGTGTGAACCAAGAACCGCATGAACCGCATGGCGCATGATCCCGAAGGGCGCGGT 238
DB 99 Ala-----GlnG1nCysSerCySValProProAlaG1y---- 109
OY 239 GGAACCAAGAGCCTTGCGAGCAGCGCAGCCCCCGGAGAGAGCCGCGCCAGGCGGC 298
DB 110 -----TyRGlInG1yAlaG1ySerVal 116
OY 299 ATGAGAGGTTCGAGCGCGAGAAATGACGCA-- 331
DB 117 LeuMetSerProValProH1sG1nMetLeuProTyMetAsnValG1yThrLeuSerArg 136
OY 332 -----CGAAGTCCGCGCAGCAAG 349
DB 137 ThrGluLeuG1nLeuLeuAengInLeuH1sCyArGrArgLySarGArgH1sArgThrI1le 156

```

```

OY 350 TTCAGCTGTTGACAGTGGAGAGCTGGAAGTGTTCGACACACTCAATACCTGAT 409
DB 157 PheThrArgGluInLeuG1uAlaLeuCluAsnLeuPheG1uInG1uThrLyS1rProAsp 176
OY 410 GTGCCCAAGAGAGGAACTTGCCGAAACTTAGGTGTGACTGAAGACAAAGTCCGGTT 469
DB 177 ValG1yThrArgG1uInLeuAlaArgLySValH1sLeuArgG1uInG1uValG1uVal 196
OY 470 TGCCTTAAGATAAAGGCGCATGTAGGCGACATCAGAGA 511
DB 197 TrpPheLySAsnArgArgAlaLyStrPArGrArgInLySarG 210

RESULT 5
A42768
homeotic protein gooseoid [similarity] - mouse
C/Species: Mus musculus (house mouse)
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 01-Sep-2000
C/Accession: A42768
R: Blum, M.; Gaunt, S.D.; Cho, K.W.Y.; Steinbeisser, H.; Blumberg, B.; Bltner, D.; De Rot
Cell 69, 1097-1106, 1992
A/Title: Gastrulation in the mouse: the role of the homeobox gene gooseoid.
A/Reference number: A42768; M01D:92315328; PMID:1352187
A/Accession: A42768
A/Status: preliminary; not compared with conceptual translation
A/Residues: 1-256 <BLU>
A/Cross-references: GB:M85271; NID:9193895; PIDN:AAA7826.1; PID:9193896
A/Note: sequence extracted from NCBI backbone (NCBI:P108110)
C/Superfamily: homeotic protein gooseoid; homeobox homology
C/Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:161-217/Domain: homeobox homology <HOX>

Alignment Scores:
Pred. No.: 3,666-08 Length: 256
Score: 187.00 Matches: 59
Percent Similarity: 39.53% Conservative: 26
Best Local Similarity: 27.44% Mismatches: 64
Query Match: 17.81% Indels: 66
DB: 1 Gaps: 6

US-09-867-753-1 (1-577) x A42768 (1-256)
OY 2 CCAACATCAGGCGCTCCAGCCATGCGCGCTGCTCCAGCAGCACCGTGTCTACTGC 61
DB 27 ProSerAlaAlaAlaProValAlaPheProAlaLeuH1sG1yAspSerLeuTyR----- 44
OY 62 CTGAGTGTATACCAAGTAATAAATTAAGCCCACTCAAGTGGGGGACAGATCAAGCGCA 121
DB 45 -----GlyAlaG1yG1yG1yThr 50
OY 122 GAAGGCCATGTTGGCCAGAGAGCTCCAGGCTCATGGGTAATGAACCTTGAGGCG-- 178
DB 51 SerSerAspTyRg1yAlaPheTyRProArGrProValAla-----ProG1yG1yAla 67
OY 179 GGTGTGAACCAAGAGAGCGCATGAACCGCATGGC----- 214
DB 68 GlyLeuProAlaAlaValG1ySerSerArgLeuG1yTyRAsnSerTyRPhyRg1yG1n 87
OY 215 -----GGCATGATCCCGCAGGCGGC 235
DB 88 LeuH1sValG1nAlaAlaAlaProValG1yProAlaCyS1yG1yAlaValProProLeuG1y 107
OY 236 GGTGAAACCAAGAGCCTTGCGAGCAGCGCAGCCCCCGGAGAGAGCCGCGCCAGGCG 295
DB 108 Ala-----GlnG1nCysSerCySValProThrProG1yTyRGlInG1yProG1ySer 125
OY 296 GCATGAGAGGTCGAGCGCGAGAAATGACGCA-- 331
DB 126 ValLeuValSerProValProH1sG1nMetLeuProTyMetAsnValG1yThrLeuSer 145
OY 332 -----CGAAGTCCGCGCAGCAAG 346

```

DB 146 ArgThGluLeuGlnLeuLeuAenGlnLeuHISCyAaRgArGlySaRgArGhISaRgThr 165
 QY 347 AAGTTCACGCTGTTGCGAGTGAGAGAGCTGGAAGTGTTCGCAACACTCAATACCT 406
 DB 166 TlePherThAspGluGlnLeuGlnAlaLeuGlnAsnLeuPheGlnGlnThrIleTyPro 185
 QY 407 GATGTCGCCCAAGAAAGGAACTTGCAGAAACTTGAAGTGTGACTGAAGCAAAAGTCCG 466
 DB 186 AspValGlyThrArgGlnGlnLeuAlaArgGlyValHISLeuArgGlnGlnValGln 205
 QY 467 GTTGTGTTTAAGAAATAAAGGCGCAGATGATGAGCATACAGCA 511
 DB 206 ValTrpPheLysAsnArgAlaLysTrpArgArgGlnLysArg 220

RESULT 6

homeobox protein Pmx - mouse
 N:Alternate names: homeotic protein K-2b
 C:Species: Mus musculus (house mouse)
 C>Date: 02-Jul-1996 #sequence revision 02-Jul-1996 #text_change 24-Sep-1999
 C:Accession: I48902; I53118; S26077
 R:Kern, M.J.; Argao, E.A.; Birkenmeier, E.H.; Rowe, L.B.; Potter, S.S.
 Genomics 19, 334-340, 1994
 A>Title: Genomic organization and chromosome localization of the murine homeobox gene Pmx
 A:Reference number: I48902; PMID:94245205; PMID:7910581
 A:Accession: I48902
 A:Status: Preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-217 <RES>
 A:Cross-references: EMBL:U03873; NID:g460124; PIDN:AA52139.1; PID:g460125
 R:Cherlyes, P.; Lilly, B.; Bryson, L.J.; Wang, Y.; Saason, D.A.; Olson, E.N.
 Development 115, 1087-1101, 1992
 A>Title: MHOX: a mesodermally restricted homeodomain protein that binds an essential site
 A:Reference number: I53118; PMID:93083424; PMID:1360403
 A:Accession: I53118
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-217 <RES>
 A:Cross-references: GB:U06902; NID:g199583; PIDN:AA39672.1; PID:g199584
 R:Kern, M.J.; Witte, D.P.; Valerius, M.T.; Aronow, B.J.; Potter, S.S.
 Nucleic Acids Res. 20, 5189-5195, 1992
 A>Title: A novel murine homeobox gene isolated by a tissue specific PCR cloning strategy
 A:Reference number: S26076; PMID:93027261; PMID:1383943
 A:Accession: S26077
 A:Molecule type: mRNA
 A:Residues: 1-217 <KER>
 A:Cross-references: EMBL:X59726
 A:Experimental source: fetal heart
 C:Genetics:
 A:Gene: Pmx
 C:Superfamily: unassigned homeobox proteins; homeobox homology
 C:Keywords: alternative splicing; DNA binding; homeobox; nucleus; transcription regulation
 F:95-151/Domain: homeobox homology <Hox>

Alignment Scores:

Pred. No.:	6e-08	Length:	217
Score:	184.50	Matches:	49
Percent Similarity:	44.97%	Conservative:	27
Best Local Similarity:	28.99%	Mismatches:	54
Query Match:	17.57%	Indels:	39
DB:	2	Gaps:	5

US-09-867-753-1 (1-577) x I48902 (1-217)

QY 113 TCAAGCGAGAAAGGCGATGTTGCGCAAGAGCTCCAGGCTCATGGGTATATGAACCT 172
 DB 2 ThnSerSerTyGlyHisValLeuGlnArgGlnProAlaLeuGlyArgGlnLeuAspSer 21
 QY 173 GAGGCGC-----GGTGAACCAAGAGGCGCGC 199
 DB 22 ProGlyAsnLeuAspThrLeuGlnAlaLysValAsnPheSerValSerHisLeuLeuAsp 41
 QY 200 ATGAACCGGCGATGCGCGCGATGATCC-----GAGGCGCGC 235

DB 42 LeuGlnGlnAlaGlyAspMetValAlaAlaGlnAlaAspGlnSerValGlyGlnAlaGly 61
 QY 236 -----GGTGAACCAAGAGGCGCGCAGACCG 265
 DB 62 ArgSerLeuLeuGlnSerProGlyLeuThrSerGlySerAspThrProGlnGlnAsp--- 80
 QY 266 CAGCCCCCGGAGAGAGCGCGCCAGCGCGCATAGAGAGGTCGCGAGCCGAGAACATG 325
 DB 81 -----AsnAspGlnLeuAsnSerGlnGlnLysValSlySlyArg 92
 QY 326 CAGCCACGAACTCGGCGCGCAAGATTACAGCTGTTGCGAGAGAGAGCTGGAAGTGT 385
 DB 93 LysGlnArgArgAsnArgThrThrPheAsnSerSerGlnLeuGlnAlaLeuGlnVal 112
 QY 386 TTCCGACACATCAATACCTCGATGCGCCACAGAGGAACTGCGCAAACTTAGGT 445
 DB 113 PheGlnArgThrHisTyTrpAspAlaPheValArgGlnAspLeuAlaArgArgValAsn 132
 QY 446 GTGACTGAAGACAAAGTGGCGGCTTTGTTTAAGAAATAAAGGCGCAGATGATGCGACAT 505
 DB 133 LeuThrGlnAlaArgValGlnValTrpPheGlnAsnArgArgAlaLysPheArgArgAsn 152
 QY 506 CAGAGAAATTAATGCTCCCAATCA 532
 DB 153 GluArg---AlaMetLeuAlaAsnLys 160

RESULT 7

homeotic protein K-2a - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 25-Feb-1994 #sequence revision 01-Sep-1995 #text_change 24-Sep-1999
 C:Accession: S26076
 R:Kern, M.J.; Witte, D.P.; Valerius, M.T.; Aronow, B.J.; Potter, S.S.
 Nucleic Acids Res. 20, 5189-5195, 1992
 A>Title: A novel murine homeobox gene isolated by a tissue specific PCR cloning strategy.
 A:Reference number: S26076; PMID:93027261; PMID:1383943
 A:Accession: S26076
 A:Molecule type: mRNA
 A:Residues: 1-245 <KER>
 A:Cross-references: EMBL:X59725; NID:g51361; PIDN:CAA42410.1; PID:g51362
 A:Experimental source: fetal heart
 C:Superfamily: unassigned homeobox proteins; homeobox homology
 C:Keywords: alternative splicing; DNA binding; homeobox; nucleus; transcription regulation
 F:95-151/Domain: homeobox homology <Hox>

Alignment Scores:

Pred. No.:	6e-08	Length:	245
Score:	184.50	Matches:	49
Percent Similarity:	44.97%	Conservative:	27
Best Local Similarity:	28.99%	Mismatches:	54
Query Match:	17.57%	Indels:	39
DB:	2	Gaps:	5

US-09-867-753-1 (1-577) x S26076 (1-245)

QY 113 TCAAGCGAGAAAGGCGATGTTGCGCAAGAGCTCCAGGCTCATGGGTATATGAACCT 172
 DB 2 ThnSerSerTyGlyHisValLeuGlnArgGlnProAlaLeuGlyArgGlnLeuAspSer 21
 QY 173 GAGGCGC-----GGTGAACCAAGAGGCGCGC 199
 DB 22 ProGlyAsnLeuAspThrLeuGlnAlaLysValAsnPheSerValSerHisLeuLeuAsp 41
 QY 200 ATGAACCGGCGATGCGCGCGATGATCC-----GAGGCGCGC 235
 DB 42 LeuGlnGlnAlaGlyAspMetValAlaAlaGlnAlaAspGlnSerValGlyGlnAlaGly 61
 QY 236 -----GGTGAACCAAGAGGCGCGCAGACCG 265
 DB 62 ArgSerLeuLeuGlnSerProGlyLeuThrSerGlySerAspThrProGlnGlnAsp--- 80
 QY 266 CAGCCCCCGGAGAGAGCGCGCCAGCGCGCATAGAGAGGTCGCGAGCCGAGAACATG 325


```

Db      81  -----Asnae6GlnIleuAensSerGIuGIuLyLylarG 92
Oy      326 CAGCCAGCAACTCGGGCGCAGCAAGTTCAAGCGTTGTCAGGTCGGAGAGCTGGAAAGGTT 385
Db      93  LysGIaIdaGArGaAsnArGIuThrPheAsnSerSerGIuLeuGIaIaLeuGIaArGVal 112
Oy      386 TTCCGACACACTCAATACCTGATGTGCCCAAGAAAGGAGCACTTGCCGAAACTTGGT 445
Db      113 PheGIuArGIuThrIhIeTyPProAspAlaPheValArGIuAerLeuAlaArGIuValAsn 132
Oy      446 GTGACTGAAGCAAAAGTCGGGTTTGTTTAAAGATTAAGGGCCAGATGTAGCCGACAT 505
Db      133 LeuThrGIuAlaArGIuValGIaIuValTTPheGIaIaSerArGIaArGIaAlaTyPheArGIaArGAsn 152
Oy      506 CAGAGAGAAATTATGTCTGCCCAATGAA 532
Db      153 GIuArG---AlaMetLeuAlaAsnLyS 160

RESULT 8
S70617      homeotic protein goosecoid - fruit fly (Drosophila melanogaster)
N:Alternate names: homeobox protein goosecoid
C:Species: Drosophila melanogaster
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 17-Nov-2000
C:Accession: S70617
R:Hahn, M.; Jaekle, H.
EMBO J. 15, 3077-3084, 1996
A:Title: Drosophila goosecoid participates in neural development but not in body axis f
A:Reference number: S70617; MUID:96272167; PMID:8670808
A:Accession: S70617
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-419 <HAA>
A:Cross-references: EMBL:U53968; NID:g1399586; PIDN:AB17948.1; PID:g1399587
C:Genetics:
A:Gene: gsc
A:Cross-references: FlyBase:FBgn0010323
A:Map position: 2
C:Function:
A:Description: plays a role in neurogenesis in post-gastrula Drosophila embryos
A>Note: not required for gastrulation like Xenopus goosecoid; expressed most strongly in
C:Superfamily: fruit fly homeobox protein goosecoid; homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:287-343/Domain: homeobox homology <Hox>

Alignment Scores:
Pred. No.: 1,21e-07 Length: 419
Score: 181.00 Matches: 48
Percent Similarity: 49.30% Conservative: 22
Best Local Similarity: 33.80% Mismatch: 42
Query Match: 17.24% Indels: 30
DB: 1 Gaps: 5

SJS-09-867-753-1 (1-577) x S70617 (1-419)
Oy      107 GCAGCACTAAACCGCAGAAAGGCGCATGGCCGCAAGAGTCCAGGCGCTCATGGCTAAT--- 163
Db      230 AlaAlaAlaGIaImetGAlaIhIaIhIaIaSerGIuAlaAlaAlaGIuLeuSerGIuIhIe 249
Oy      164 -----ATGAACCTTAGAGGCGCGTGTGAACCAAGCAAGAAC---GGCATGAACCGCGAT 211
Db      250 HIsHtRPkHtHIsRPkHtHIsRPkHtHIsRPkHtHIsRPkHtHIsRPkHtHIsRPkHtHIs 269
Oy      212 GGC---GGCATGAATCCCGAGAGGCGCGGTGTGAACCAAGCAAGAGCTCGGAGAGCCGCGAG 268
Db      270 GIuGIaIhIaIhIaIhIaIhIaIhIaIhIaIhIaIhIaIhIaIhIaIhIaIhIaIhIa 280
Oy      269 CCCCCCGCGAGAGCGAGCCGCGAGCGCGCATGAGAGGTCGCGAGCCGAGAAACATGACAG 328
Db      281 ProProProLyS-----Arg 285

329 CCAAGAACTCGGCGCGCAAGAGTTCAAGCTGTTCGAGGTGAGAGACTGGAAAGGTGTTTC 388

```

RESULT 9

JM0097
bicoid-related homeobox protein Ptx2 - chicken
C|Species: Gallus gallus (chicken)
C|Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 15-Oct-1999
C|Accession: JM0097
R|Biochem. Biophys. Res. Commun. 247, 100-105, 1998
A|Title: Cloning and expression pattern of chicken Ptx2: A new component in the SHN sig
A|Reference number: JM0097; MUID:98300291; PMID:9636662
A|Accession: JM0097
A|Molecule type: mRNA
A|Residues: 1-333 <STA>
C|Comment: This protein is a new component in the Sonic hedgehog signaling pathway and p)

C|Gene: cPtx2
C|Superfamily: chicken bicoid-related homeobox protein Ptx2; homeobox homology
C|Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:102-158/Domain: homeobox homology <Hox>

Alignment Scores:

Pred. No.:	1,47e-07	Length:	333
Score:	180.00	Matches:	48
Percent Similarity:	45.56%	Conservative:	29
Best Local Similarity:	28.40%	Mismatches:	40
Query Match:	17.14%	Indels:	52
DB:	1	Gaps:	5

US-09-867-753-1 (1-577) x JM0097 (1-333)

Dp 286 LysARGNIARgThrllePherthrGlugInLeuGluInLeuGluAlatThrPe 305
Dy 389 CGACACTCATATACCCTGGATGTGCCACAGAGGAACTTGCCGAACCTTAGCGTGTG 448
Db 306 AsplvrtlnHtstYrProAsprValVallLeuArgGlugInLeuAlatLeuLysValAspleu 325
Dy 449 ACTGAACAAGAAGTCGGGTTGTGTTTAAGATTAAAGGCCGATGTAGCCGACATCAC 508
Db 326 LysGIUGIUAArgValGLuValTrlPhelybaANArGARAlalAtylstrPAglYsgInLys 345
Dy 509 AGAGAA 514
Db 346 ArgGlu 347

Dp 5 ACATCAGGCGCTCCAGCCATGCGCGCTTGCTGCATCCAGACACCGTGTACTGCGCTG 64
Dy 45 ThrAlatEuAlarProGlyGlnAlatrgSerSerleu---GlnAlalAlayshIARgLeu 63
Dy 65 AGTGTATATCCAGGTAAATAAAGCCCCACACTCAGCTGGGGCAGCATCAAGCGAGAA 124
Db 64 GluValIHtsthrlleSeraSprThrSerSerProGlu-----AlaAlaglu 78
Dy 125 GAGCATGTTGGCCAAGAGAGCTCAGGCGCTCATGCGGTATATGAACCTTAGGGCGGTGTG 184
Db 79 LysGIULysSerGln----- 83
Dy 185 AACCAAGAGAACCGCATGAACCGCATGCGGCATGATCCCGAGGGCGCGGTGAAC 244
Db 84 -----GlnGlyLysSerGluAspAlagly----- 91
Dy 245 CAGAGAGCTCGGCACAGCGCCAGGCCCGCGGAGAGAGCGGCCGACATGAG 304
Db 92 -----ProGluAspProser----- 96
Dy 305 GGTCGACGCCGAGAACATYGACGACCGAACTCGGCGGACGAAAGTTCAAGCTGTGACG 364
Db 97 -----LysLysLysArgGlnArgGlnArgGlnArgThrsIspheThrSerGlnGln 112
Dy 365 GTGAGAGAGCTGGAAAGTCTTTCGACACACACTCAATACCTCGATGTGCCCAAGAAGG 424
Db 113 LeuGlnGlnLeuGlnAlatThrPhelGlnArgAsnArgTyrProAspMetSerTrlrArgGlu 132

